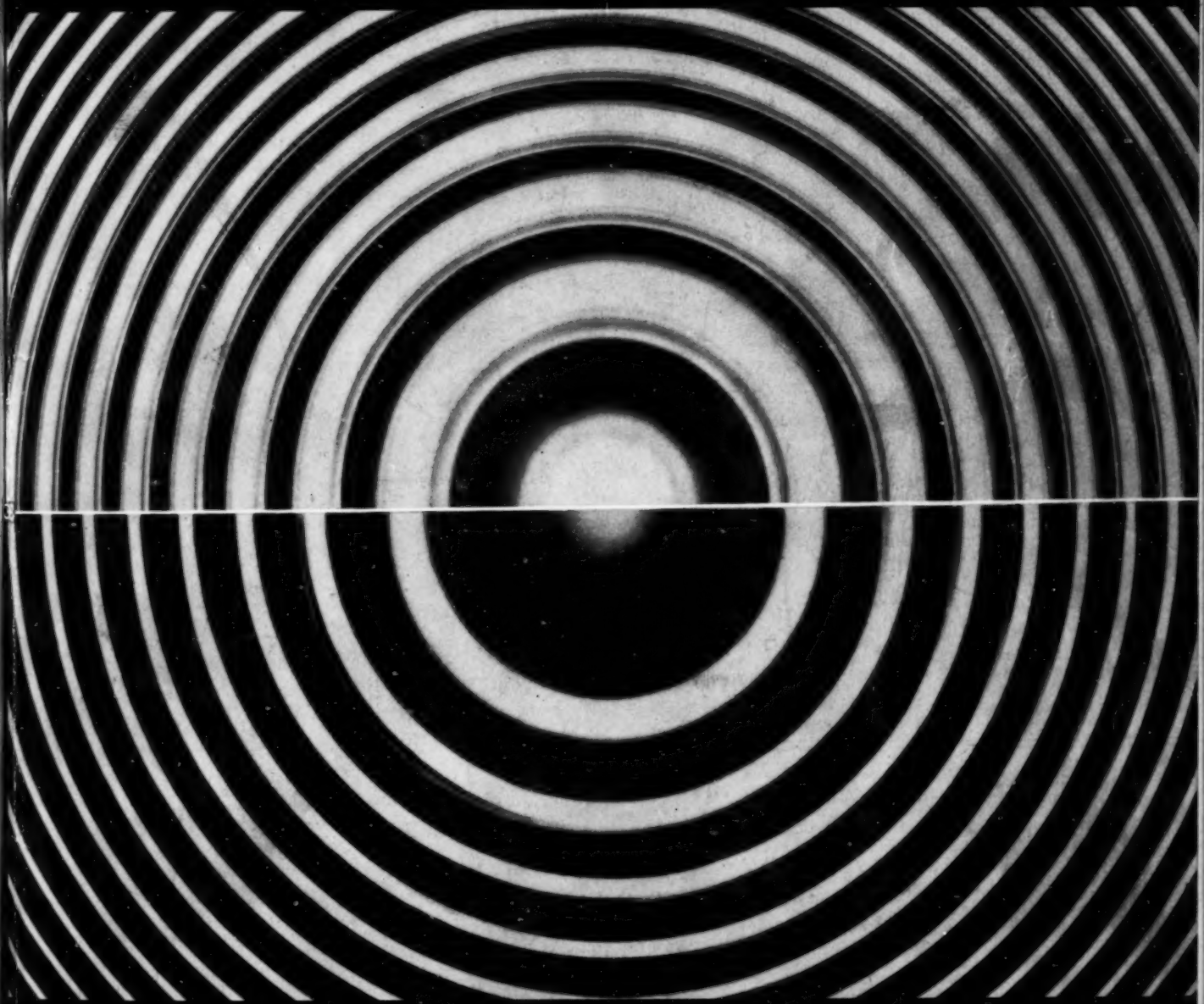
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 19, No. 11 NOVEMBER, 1948

In This Issue

		Page
Contribution of Modern Physics to Metallurgy	FREDERICK SEITZ	973

Contributed Original Research

Velocity of the Anode Spark in Copper Sulfate Solutions Under Application of Impulsive Potential	HUGH F. HENRY	988
Bidwell's Intercept Relation and the Thermal Conductivity of Liquid Metals	R. W. POWELL	995
Effective "Gamma" for Isentropic Expansions of Real Gases	ARTHUR S. IBERALL	997
On the Mixing Properties of Non-Linear Condensers	A. VAN DER ZIEL	999
Extension of Lagrange's Equations to Electromagnetic Field Problems	PRESCOTT D. CROUT	1007
Observation on Heat Sources Placed at Random on a Surface	GREGORY H. WANNIER	1020
Wave Guides for Slow Waves	L. BRILLOUIN	1023
Divergent Beam X-Ray Photography with Standard Diffraction Equipment	A. H. GEISLER, J. K. HILL, AND J. B. NEWKIRK	1041
Ion Microscope with a Transverse Magnetic Field	NOBUJI SASAKI	1050
Growth of Charged Particles in Clouds	WULF KUNKEL	1053
Magnitude and Character of Errors Produced by Shape Factors in Stokes' Law Estimates of Particle Radius	WULF B. KUNKEL	1056
On the Measurement of Cavity Impedance	W. W. HANSEN AND R. F. POST	1059
Fracture of Liquids	JOHN C. FISHER	1062
Geometrical Factors Affecting the Contours of X-Ray Spectrometer Maxima. I. Factors Causing Asymmetry	LEROY ALEXANDER	1068
Theory of the Positive Column in an Annular Discharge Tube	JAROSLAV PACHNER	1072
Electron Diffraction Study of Particle Size in Thin Bi Films Deposited by Evaporation in Vacuum	C. T. KEOGH AND A. H. WEBER	1077
Plastic Flow, Creep, and Stress Relaxation. Part IV. Anomalous Flow as an Order-Disorder Transition	CHARLES MACK	1082

Letters to the Editor:

Note on "Statistical Aspects of Fracture Problems"	FRANKLIN H. FOWLER, JR.	1092
Transmission Properties of Materials in the Millimeter Wave Region	E. WANTUCH, J. H. ROHRBAUGH, J. P. COOLEY, J. H. GREIG, AND J. SHMOYS	1092
Some Remarks Concerning a Proposed Method for Testing the Brittleness of Elastomers	C. H. KLUTE	1093
On the Flow of a Perfect Fluid between Two Plates from a Circular Inlet Channel	E. T. BENEDIKT	1094

In Every Issue

Calendar of Meetings	1055
Here and There	1095
New Booklets	1095
Index to Advertisers	xxxvi

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

Associate Editors

R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE
W. JAMES LYONS
PHILIP MORRISON
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND
JOHN VON NEUMANN

Publication of the

American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*

CLEVELAND NORCROSS, *Assistant Director*

RUTH F. BRYANS, *Publication Manager*

THEODORE VORBURGER, *Business Manager*

JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD

GEORGE R. HARRISON, *Chairman*

WALLACE WATERFALL, *Secretary*

GEORGE B. PEGRAM, *Treasurer*

LUIS W. ALVAREZ
ROBERT B. BRODE
E. U. CONDON
A. J. DEMPSTER
W. F. FAIR, JR.
HARVEY FLETCHER
R. C. GIBBS
W. V. HOUSTON
PAUL E. KLOPSTEG
F. W. LOOMIS
PHILIP M. MORSE
A. H. PFUND
JOHN T. TATE
M. W. WHITE
A. G. WORTHING

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. AND CANADA	ELSEWHERE
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 18, 1947—$132.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1948, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 19, Number 11 November, 1948

The Contribution of Modern Physics to Metallurgy*

FREDERICK SEITZ
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received May 17, 1948)

I. INTRODUCTION

PHYSICS is primarily a branch of pure science which has developed because men possess a deep-seated curiosity concerning the internal operation of the world about them. In this sense physics has much in common with philosophy; in fact physics was called "Natural Philosophy" among English-speaking people in the recent past. Since almost all knowledge concerning the behavior of nature ultimately proves to be of use in the processes of practical living, the various branches of physics have yielded from time to time rich harvests of useful information.

The contributions of physics to technology usually take either of two forms: In certain cases developments in physics have opened entirely new fields of technology—fields which probably would have been left completely closed if it had not been for the fundamental investigations which were based purely upon scientific curiosity. For example, the fields of electrical engineering and of atomic power development lie in this category. In both cases science is unambiguously the parent of practical development in the broad sense. In other cases the field of technology was opened and established before the natural laws which operate in the field received scientific attention. Subsequent scientific work aided in the advance of the field either by providing a clarification of the fundamental principles which were used without an explicit understanding of their existence and implications, or by indicating ways in which the field could be expanded or joined with other branches of technology. Mechanical engineering is an excellent example of a branch of technology in which much of the early development was carried out without explicit use of the underlying natural laws. However, subsequent scientific discovery added greatly to the usefulness of the field. Hydraulics and ballistics were well developed branches of mechanical engineering before the laws of mechanics were discovered by Newton and his contemporaries; heat engines had been in operation for over a century before the laws of thermodynamics were clearly understood.

Metallurgy, like mechanical engineering, is a field which reaches so far into the past that much of the early practical advances were made without the aid of a highly developed scientific background. To appreciate this fact it is only necessary to realize that bronze, brass, and steel of good quality were made in prehistoric times. The discoveries of science have pervaded the various branches of metallurgy gradually and at a nonuniform rate.

* This manuscript represents one of the Gordon McKay Lectures in Applied Science delivered at Harvard University in 1947.

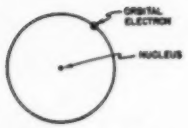


FIG. 1. Schematic diagram showing the relation between the orbital electrons and the nucleus of the atom. The diameter of the nucleus, which contains most of the mass, is about one hundred thousand times smaller than the diameter of the orbit of a typical valence electron.

Fields of technology in which practical advances precede fundamental understanding in a substantial way exhibit certain characteristic features which are not as clearly evident in fields which have sprung nearly full-grown from science. In all such cases the first important discoveries are made by diligent application of trial and error methods which depend far more upon patience, keen observation and manual skill than upon basic understanding. Once a successful method of achieving a given result is obtained, it is passed on from master to apprentice more as a matter of art than of science; in short the traditions of the field come to resemble those of the artist rather than the scientist. When the scientific background of the field is eventually established, an opportunity arises to alter the traditions to take advantage of the newer developments. The speed with which this occurs depends both upon the firmness with which the old traditions are established and upon the practical values which accompany the new scientific discoveries. Generally speaking the better established parts of the field yield more slowly than those which are in the process of forming. Steel making, which has a very old background, was affected less rapidly by modern discoveries in physics and chemistry than the fields of nonferrous metallurgy which have been developed in the past few decades. However slow the transition in certain branches of metallurgy may be, the end result seems to be unambiguous: The basic traditions of leading metallurgical research will eventually become those of the physicist and chemist even though the methods of manufacture will undoubtedly exhibit the influence of early history for a long time to come.

The importance of modern physics in the present development of metallurgy rests primarily upon two facts: Metals are composed of atoms and modern physics is concerned with obtaining a deeper understanding of the properties of atoms. Each advance in understanding of atomic structure has made possible a deeper understanding of the factors which determine the characteristic properties of aggregates of atoms. In this sense the developments of modern physics have made it possible to understand the laws of classical valence chemistry, which are applicable to most organic compounds and to the simpler inorganic ones, and to extend these laws to cover cases in which the older forms are not valid, most particularly to metals. This extension is accompanied with a deeper understanding of many of the most characteristic properties of metals, such as their high electrical conductivity and ductility. Attention will be focused upon this aspect of the application of modern physics to metallurgy in the present lecture.

It is to be emphasized that the contributions to metallurgy from chemistry and from the instrumental developments of physics occupy a position comparable to those from atomic physics. Analyses of these however are fitting subjects for other lectures.

II. ATOMIC STRUCTURE

It is an underlying principle of modern science that all matter is composed of atoms which retain their identity unchanged in all ordinary chemical reactions. Explorations of atomic structure early in this century showed that the atom has many characteristics in common with a minature solar system. Most of the mass is concentrated in a very small central core or *nucleus* which is positively charged and which is now known to be composed of positively charged protons (the nuclei of ordinary hydrogen) and neutrons, which are uncharged counterparts of protons. About this nucleus swarm negatively charged electrons, usually called the *orbital electrons*, which are very light in comparison with the nucleus (Fig. 1). The electronic charge has the same magnitude as the proton charge, but is opposite in sign. In the normal neutral atom the number of electrons swarming about the nucleus is equal to the number of protons in the nucleus so that the charges balance. The chemical properties of the atom are determined by the number of orbital electrons; neutral atoms having identical numbers of orbital

electrons have essentially identical chemical properties. This means, in turn, that the chemical properties are specified by the number of protons in the nucleus, a quantity that is called the *atomic number*. Not all atoms having the same atomic number possess the same number of neutrons in the nucleus; hence atoms of the same chemical species do not necessarily possess the same mass.

In ordinary chemical reactions only the outermost orbital electrons are appreciably affected. The manner in which they are affected is of course of paramount importance for the theory of chemical binding. In salts, such as sodium chloride, electrons are transferred from the metallic element to the non-metallic one and the atoms become charged, or are transformed to ions. In most organic compounds and in practically all metals the atoms are not *ionized* to the same degree as in salts. Instead chemical binding takes place as a result of the shearing of electrons between atoms on a more nearly equitable basis.

The distance by which atoms are spaced in ordinary solid or liquid matter is determined by the dimensions over which the orbital electrons swarm, that is by the diameter of the orbit of the outermost electrons in the atom or ion. These diameters are remarkably constant for a given atom or ion in a number of different chemical compounds so that to a first approximation the atoms or ions can be regarded as rigid spheres which are in tangential contact (Fig. 2). It is to be emphasized that the diameter varies when electrons are added to or subtracted from the atom and hence is related to the state of ionization. The diameter of an atom or ion can be altered by subjecting the compound in which it is bound to high pressures, so that the rigid-sphere concept of the atom must be used discriminately.

The problem of determining the laws of motion obeyed by the orbital electrons in atoms occupied physicists for the first thirty years of this century. The results of this study, which is one of the most brilliant in the history of science, is summarized in the subject of wave mechanics which is believed to be a rigorous representation of natural law. In principle at least it is possible to determine the exact behavior of the electrons in any chemical compound by application of this branch of mechanics. Unfortunately the equations of motion are sufficiently complicated that most problems of practical interest can be treated only by approximate methods at the present time. For this reason the practice of applying wave mechanics has become sufficiently specialized that it is unsuited for presentation in a lecture of this type. Fortunately, the results of the theory are not difficult to visualize and may be presented without reference to the mathematical techniques of the theory.

From our standpoint the most important results of the theory are as follows:

(1) All electrons possess wave properties in addition to the corpuscular characteristics which we commonly associate with small particles of matter. This means that an electron in motion has associated with it a vibrational frequency and a wave-length, just as an ordinary sound wave or light wave does. The wave-length λ is simply related to the velocity of motion by the equation

$$\lambda = h/mv,$$

in which v is the velocity of the electron, m is its mass, and h is a universal constant named in honor of the physicist Planck.

(2) The electron retains its wave.characteristics when it is bound to an atom, taking up a characteristic vibrational pattern. An important result of this is the fact that an orbital electron in an atom cannot assume an orbit possessing an arbitrary energy, but is restricted to discrete values which may be determined by solving the equations of the theory. This discreteness of energy states is an inherent characteristic of wave mechanics and differentiates this branch of mechanics from classical, or Newtonian mechanics in which continuous energy states are allowed. The relation between the discreteness of the allowed energy states of electrons and the wave

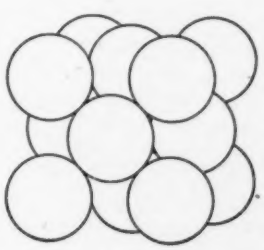


FIG. 2. A typical metal lattice (face-centered cubic type). The atoms are schematical represented by spheres which are in contact.

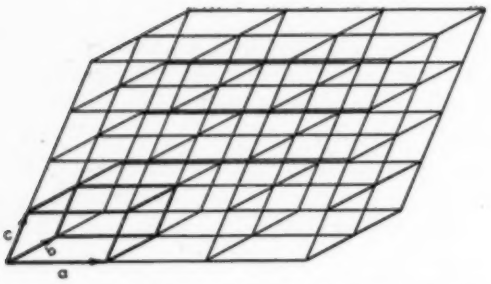


FIG. 3. A three dimensional lattice of cells. Any cell may be translated into any other by a translation involving integer multiples of the three fundamental translations a, b, and c shown in the lower left-hand corner. The atoms of the crystal are arranged in identical ways in each cell.

characteristics resembles closely the relation between the discrete vibrational frequencies of a stretched string, such as a violin string, and the wave characteristics of such vibrations.

(3) In addition to possessing wave characteristics, the electron possesses a spin, that is, it behaves as if it were rotating about an axis which passes through its center. As a result of this spin the electron possesses the magnetic attributes of a small bar magnet. This property is exceedingly important for understanding the magnetic properties of matter, particularly the ferromagnetism exhibited by iron and its alloys.

(4) Only two electrons in a single system such as an atom or cluster of atoms may have the same wave pattern and these two electrons must have opposite spin. This *exclusion principle*, discovered by Pauli, has a very far-reaching effect on the behavior of atoms since it prevents all atoms from having almost identical chemical properties. As we shall see later, it also prevents all solids from having the characteristics of metals.

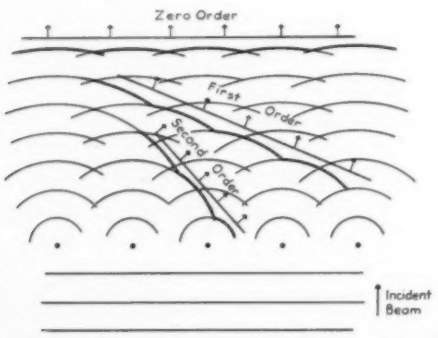


FIG. 4. The diffraction of a wave by lattice. The incident wave, whose nodes are represented by the horizontal lines at the bottom of the figure is scattered by the line of atoms, whose centers are represented by black dots. The scattered waves are spherical and are concentric with each of the atoms. The waves scattered by different atoms interfere constructively and destructively to form plane waves moving in different directions as shown.

III. STRUCTURE

All metals are composed of crystals or grains. In most metals which are used in practical service these crystals are small, being either barely visible to the naked eye or resolved only with the use of a microscope. Special techniques have been developed for obtaining large single crystals of most metals and a study of these has contributed a great deal to our present understanding of the properties of metals since in such studies it is possible to eliminate the influence of the junction between grains.

One of the very great contributions of modern physics to metallurgy has been the development of methods for determining the arrangement of atoms in individual grains. The most widely used of these methods is based on the fact that in crystals the atoms are arranged in a lattice-like array (Fig. 3) and that such a lattice will scatter or *diffract* any form of wave motion having the proper wave-length in a very regular manner (Fig. 4). Through good fortune beams of x-rays and cathode rays (electrons) can be produced which possess the proper wave characteristics to be diffracted by the lattices of most metals. An analysis of the scattered radiation, or the *diffraction pattern*, allows one to determine the arrangement of atoms on the crystals which are responsible for the scattering. If the specimen being studied is a single crystal the lattice arrangement is the principal information that is obtained; however, if the specimen is a polycrystal, additional information concerning the relative orientation of the grains can be obtained. In general it is easier to determine the lattice arrangement from studies with single crystals than with polycrystals.

The determination of the structures of metals has revealed a number of interesting and important facts of which the following are the most outstanding:

(1) The great majority of pure metals containing only one type of element possess very simple lattice structures which resemble closely the lattice configurations adopted by spheres in close

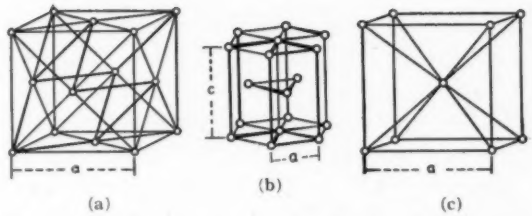


FIG. 5. (a) Face-centered cubic lattice. (b) Hexagonal close-packed lattice. (c) Body-centered cubic lattice. The first two types represent close-packed arrangements for spheres; the third possesses only slightly lower density if the spheres are in contact along the body diagonal.

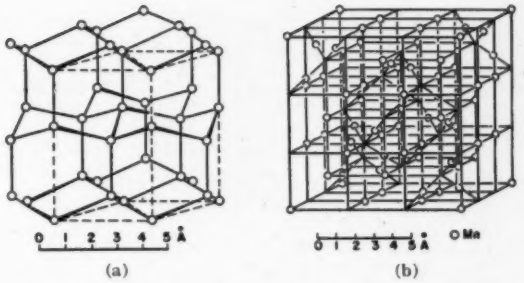


FIG. 6. (a) The lattice of bismuth. (b) The lattice of one of the phases of manganese.

packing. The three most common structures are shown in Fig. 5. The first two of these actually are arrangements for the densest packing of spheres; the third possesses only slightly lower density. A few metals, such as bismuth and manganese have more complicated structures (Fig. 6) however, these cases represent striking exceptions.

Once the lattice arrangement has been determined, it is clear that the atomic diameter is given immediately as the spacing between the centers of the nearest neighboring atoms. In this way values of the diameters of all atoms and ions have been determined. Values for the metallic elements are shown in Fig. 7.

(2) The compounds of metals, namely alloys, are distinguished chemically by the fact that the constituent elements can combine continuously for wide ranges of composition, in contrast with the situation for salts and organic compounds. For example zinc can be added to copper continuously in amounts varying from zero to forty percent without changing the basic lattice structure. Larger amounts of zinc can be added to copper with a change in structure and each such structure or *phase* can exist for a range of composition (Fig. 8).

Structural analysis shows that alloys can be divided into two distinct categories namely *substitutional* and *interstitial*. In the former, the combining atoms have comparable sizes and as the composition is varied continuously within a given phase the atoms of the one or more constituent which are increased replace those of the constituents which are diminished. Thus to a high degree of approximation, the volume of the alloy is the sum of the volumes of the constituents when considered separately as pure metallic elements. In interstitial alloys one of the constituents is considerably smaller than the other and occupied interstitial positions in the lattice of the larger atoms. In such cases the fraction of interstitial lattice sites occupied by the smaller atom increases as the fraction of interstitial constituent increases.

A good example of an interstitial alloy is palladium which has been exposed to hydrogen and has had an opportunity to absorb the gas. The hydrogen atoms, which are very small, occupy the interstitial positions in the face-centered cubic lattice of palladium. Certain phases of steel, the alloy of carbon and iron, also represent good examples of interstitial alloys. In this case carbon is the interstitial element. For example, Austenite, a phase which is stable above 721°C, is a face-centered cubic lattice of iron atoms in which carbon atoms may be dissolved in a range from zero to 8 percent. Some of the most valuable

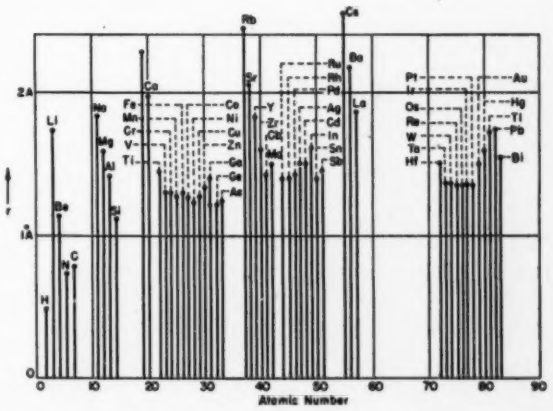


FIG. 7. The radii of the atoms. The values are obtained by taking half the spacing between the centers of nearest neighboring atoms in the lattices of the pure elements. Two or more values are given in cases in which two or more lattices are known and the values for the different lattices differ substantially.

properties of steel are derived from the fact that this phase decomposes below 721°C. Steel is hardened by controlling this decomposition properly.

A careful study of atomic diameters shows that substitutional alloys form readily only when the diameters of the constituents are the same to within fifteen percent. Similarly, interstitial alloys form readily only when the diameter of the smaller atom, which occupies the interstitial positions in the lattice of the larger, is less than 0.60 times the diameter of the larger atom. In other words these alloys form readily only when the lattice is not disturbed too violently by the presence of misfit atoms. Metals whose atomic diameters differ by more than fifteen percent and less than forty percent do not combine readily and usually do not lead to useful products when they do form. It is interesting to note that steel is a borderline case in which the diameter of the carbon atom almost exceeds the limiting value. This fact undoubtedly explains the absence of an ideal interstitial alloy at temperatures below 721° and provides steel with some of its most valuable properties, namely those related to the precipitation of iron-carbide below 721°C.

(3) Substitutional alloys, when examined as a whole, show remarkable regularity. It is found, for example, that similar successions of phases are found in very different systems as the composition is varied. Thus the copper-zinc system is very similar to the copper-tin system even though zinc and tin are very different chemically in the sense that zinc is divalent and tin is quadrivalent. Hume-Rothery has made a careful study of this type of regularity and has shown that in general the properties of an alloy are determined by the ratio of the number of valence electrons to the number of atoms. Thus the copper-zinc and copper-tin phase diagrams become practically superposable, particularly at the copper-rich end of the diagrams, if the electron-atom ratio is used on the horizontal axis instead of the atomic composition. This principle indicates that the valence electrons determine the structure of these alloys in a systematic way, in analogy with the situation for inorganic and organic compounds even though the "rules of combination" are very different in the two cases.

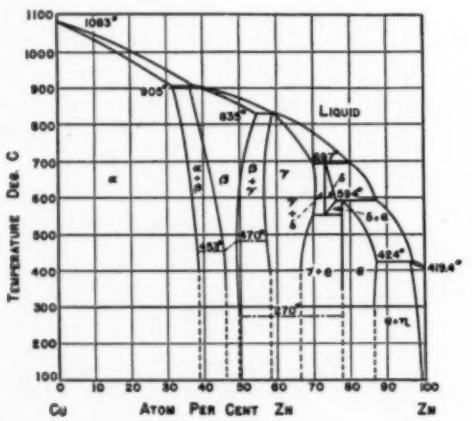


FIG. 8. The phase diagram of the copper-zinc, or brass, system. The α phase is face-centered cubic; the β phase is body-centered cubic and exhibits ordering below 470°C; the γ phase has a complex structure possessing 52 atoms per unit cell.

(4) Certain substitutional alloys, of which the body centered cubic phase of brass having approximately 50-50 atomic percentages of copper and zinc is an example, exhibit an interesting structural change which has been widely studied. At elevated temperatures (above 470°C), the copper and zinc atoms are randomly distributed throughout the lattice; however, below 470°, the zinc atoms prefer body centered positions, whereas the copper atoms prefer the positions at the corners of the cubic lattice. In other words, the over-all lattice structure is not affected by the transition but the copper and zinc atoms take up preferred positions within the lattice. It has been shown by Mott that the force determining ordering at the lower temperature can be related to relatively slight differences in charge gained by the atoms in the solid solution. Thus zinc, which is electropositive relative to copper, gains a slight positive charge whereas copper receives the complementary negative charge. In this sense the ordered alloy resembles a salt such as sodium chloride or, more exactly, cesium chloride, since the latter has the same lattice structure as the ordered brass. The charges on the metal atoms in the alloy are much smaller than the corresponding charges of the ions in salts, being about one-tenth as large. Nevertheless, even the small charge is sufficient to impart salt-like characteristics to the alloy without affecting the metallic properties in a very substantial way.

Studies of the manner in which alloys scatter x-rays and electrons have yielded a great deal of

useful information beyond the determination of lattice structures and the quantities, such as atomic diameters, that may be obtained from such structures. These lattices are often not perfect and a careful investigation of the scattering pattern may yield very useful information concerning the nature of the imperfections. Imperfections may have many origins: They may result from mechanical working of the metal or from lack of uniformity of the concentration of the alloying agents. Since such irregularities have a profound effect upon the properties of metals it is important to obtain information concerning them in order to provide better understanding and control. This aspect of diffraction study is by far the most important from the standpoint of frontier research at the present time and is one of the most interesting fields of the physics of metals. For example, duraluminum which is a hardenable alloy of aluminum, contains several percent of copper which is in perfect solution at high temperatures, but which precipitates near room temperature. The alloy hardens during this precipitation so that this property is of far-reaching interest. Studies of the diffraction of x-rays by the alloy have yielded much important information concerning the precipitation process. Similar valuable information has been obtained regarding the distortion which accompanies the flow of metals under stress.

IV. THE ELECTRICAL CONDUCTIVITY OF METALS

The high electrical and thermal conductivity of the common metals is one of their most distinguishing characteristics and one that has excited interest for a long period of time. At the beginning of the century Drude suggested that the valence electrons of the metallic atoms become free as the result of the formation of the interatomic bond so that they are able to wander throughout the metal. According to Drude's picture, this motion is marked by collisions with the atoms of the lattice which cause deviations from a straight-line path in the same way that the molecules of a gas are deflected by collisions. An electric field will produce a directed component of motion upon the random one and produce a current. This qualitative picture has withstood the test of time, although Drude's attempts to calculate the electrical resistance quantitatively on the basis of the free electron picture proved to be ineffective. For example, experiment shows that the resistance of most pure metals drops continuously to very low values as the temperature is decreased toward the absolute zero. Drude's picture of the conduction process did not allow a clear understanding of this universal behavior. Similarly, Drude's picture implies that the free electrons should have a heat capacity that is about 100 times larger than the observed value since each electron should have the same thermal energy of motion as the atoms in a gas.

The difficulties in Drude's formulation of the free electron picture were not resolved until the development of wave mechanics, for it turned out that it was just those aspects of the behavior of electrons which can only be treated on the basis of this branch of mechanics that were responsible for the deviations between his work and experiment. The equations of wave mechanics show that if the lattice of the metal is perfect and if all of the atoms are at rest, free electrons would move undisturbed through the crystal. This would correspond to a condition in which the metal would possess vanishing electrical resistance. The situation for passage of the electron waves in this case resembles closely that observed in the passage of light waves through a transparent window pane, or of sound waves through an empty hall. In the case of pure metals the source of scattering which produces a non-vanishing electrical resistance is the thermal oscillations of the atoms. At any finite temperature the atoms of a crystal are subject to thermal agitation which produces fluctuating electric fields that scatter the electron wave in much the same way that fog droplets in an atmospheric cloud scatter light. This thermal agitation decreases very rapidly near the absolute zero of temperature and causes a corresponding drop in the electrical resistance. Impurity or alloying atoms will also have the effect of scattering the electron waves and thereby contribute to the resistance. This component of resistance does not depend appreciably upon the temperature, so that it can be observed even at the absolute zero of temperature.

The fact that the free electron gas does not possess as much energy of thermal agitation as

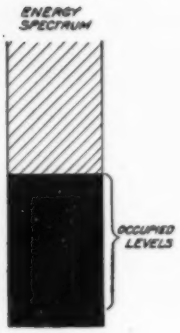


FIG. 9. The energy spectrum of the states in a metal. The levels are very closely spaced (quasi-continuous). As a result of the exclusion principle, the electrons occupy a nearly continuous portion of the spectrum very densely. The uppermost occupied levels are very close to the lowest vacant levels, as shown.

might be expected from the corresponding behavior of an ordinary gas also receives a straightforward explanation in terms of the principles of wave mechanics. The allowed states of motion of the free electrons possess discrete energy levels in accordance with the principles discussed in Section II. The allowed wave patterns of the electrons are specified in much the same way as the allowed vibrational patterns of a stretched drumhead. Only those waves which have nodes at the boundary of the specimen are permitted. Moreover, only two electrons having opposite spin are allowed in each mode of vibration so that the total aggregate of electrons in the whole specimen of metal is compelled to occupy a broad spectrum of energy states (Fig. 9). If it were not for the Pauli exclusion principle, which restricts the number of electrons which may occupy a given vibrational state, the occupied spectrum of energy states would be very narrow. In fact at the absolute zero of temperature all would occupy the state of lowest energy and a broader range would be occupied at higher temperatures only because of thermal agitation. Hence the Pauli principle has two important effects: It raises the average energy of the electrons and makes it necessary for most of the electrons to receive a large amount of energy before they can be excited to the unoccupied levels which lie above the occupied ones. Only the electrons at the very top of the distribution can receive appreciable thermal excitation. As a result the electron gas as a whole absorbs far less thermal energy that the corresponding atomic or molecular gas does.

Once the factors that influence the scattering of electron waves are clearly understood, it proves possible to use the measurement of this conductivity as a tool in determining changes that take place in the state of a metal as it is subject to various influences. For example, changes in the resistance of duraluminum that occur during the precipitation of copper yield information concerning the precipitate. Similarly, changes in the degree of order in an alloy, such as beta-brass, which shows the order-disorder phenomenon can be studied with the aid of resistance measurements.

It is to be emphasized that our present understanding of the electrical conductivity of metals is not sufficiently broad to furnish us with a clear understanding of the phenomenon of superconductivity. Certain metals, such as lead and tin, exhibit a very abrupt drop in electrical conductivity in the vicinity of the absolute zero of temperature. The electrical resistance is zero for all practical purposes below the transition temperature. This change in state is not accompanied by an observable change in structure, but is associated with abrupt changes in magnetic properties that undoubtedly have deep implications concerning the transition. Although superconductivity has found only limited practical use, it is possible that more important uses will be discovered as our understanding develops. This field can be regarded as one of the important ones in frontier research in the physics of metals.

V. THE DIFFERENCES BETWEEN METALS AND NON-METALS

If we grant that many of the properties of metals can be explained readily on the basis of wave mechanics and the assumption that they contain a gas of free electrons, we are lead to the following fundamental question: Why are the valence electrons not free in other solids, such as salts and plastics? This question evidently probes very deeply into the nature of chemical binding and a successful answer should furnish us a clearer understanding of more of the characteristic properties of metals, such as their ductility. Great progress was made in exploring this topic in the period between 1927 and 1939, and as might be supposed from the preceding discussion, wave mechanics played a decisive role in the matter.

The final conclusion drawn from the study of the wave properties of the valence electrons in solids is as follows: The valence electrons in all

tightly bound solids, such as salts, the abrasive materials (e.g., carborundum and diamond) and some of the more stable plastics, have almost the same ability to roam about the lattice as do the electrons in metals. The spectrum of allowed energy states is discrete and a range of the spectrum from the lowest level upwards is occupied in accordance with the Pauli principle, just as in metals. The difference between the two cases rests upon the fact that the highest occupied level and the lowest occupied one are very close in metals but are relatively widely separated in insulators (Fig. 10). This separation determines the ease with which it is possible to change the state of motion of the electron gas with an electric field: Such a change is easy in metals for which the spacing is small; that is the electrons occupying levels near the top of the spectrum are easy to accelerate. Conversely a change in state is relatively difficult for insulators because of the large spacing and it is difficult to induce a current.

Thus the question of whether or not a given solid is a metallic conductor or not hinges upon the distribution of the energy states and the manner in which they are occupied as a consequence of the exclusion principle. Careful investigation of this distribution on the basis of wave mechanics shows that it is determined by the lattice structure and by the distribution of electric field within the atoms which comprise the crystal. This investigation also shows that the well-known insulators and metals have exactly the types of distribution of levels that experience states must be the case. The theory also tells us however that we must not be surprised if there are metals which exhibit properties intermediate between the ideal metals and insulators. These are cases in which the density of levels at the top of the occupied region is very low, as if an incipient gap were in process of formation. In fact metals such as bismuth graphite, and certain alloys with relatively poor electronic conductivity lie in this category. Similarly we must not be surprised if there are non-metals which exhibit some of the characteristic of metals. These are cases in which the gap between the filled and empty levels are small relative to the value for the best insulators. Materials such as silicon and germanium, which have a metallic luster but are insulating at very low temperatures, fall in this category. The gap is sufficiently small in these cases that electrons become excited from the occupied to the empty levels as a result of thermal agitation. Hence the materials become better and better electronic conductors as the temperature is raised.

It is possible to vary the electronic conductivity of certain insulators by varying the composition slightly. For example, cuprous oxide having the exact composition Cu_2O is not a good electronic conductor. On the other hand, if the oxide is prepared in such a way that it contains a slight excess of oxygen, it can be made a relatively good electronic conductor, the conductivity increasing with the excess of oxygen. In such a case the excess of oxygen introduces new energy levels in the gap that exists in the ideal compound and furnishes a partial "bridge" between the insulating and metallic state. Materials which possess this ability to conduct moderately well when properly prepared in spite of the existence of a gap at the top of the occupied levels, are termed "semi-conductors." In general their resistance decreases with increasing temperatures, in contrast with metals, because thermal excitation plays an important role in exciting electrons from the occupied levels. Silicon and germanium can be regarded as if in the same category.

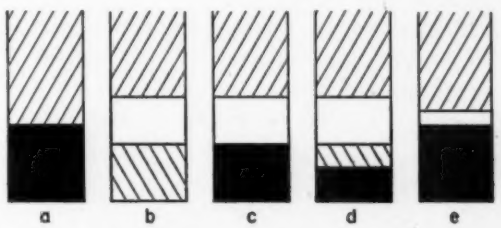


FIG. 10. The distribution of energy levels in different solid materials. Occupied levels are indicated by blackening. (a) shows a typical metal in which the levels are entirely quasi-continuous above a certain lowest value and the uppermost occupied level is very close to vacant levels. (b) shows the level spectrum found in all insulators in which a gap occurs in the quasi-continuous levels. If the occupied levels extend just to the edge of the gap. (c) the solid is an insulator since a strong electric field is needed to raise electrons to the vacant levels and induce a current. If the lowest band in only partly filled (d) the solid is a metal. Solids such as silicon and germanium which are semi-conductors even in the pure state correspond to case (e) in which the gap is very small so that electrons can be thermally excited across it at temperatures near room temperature.

VI. THE SOURCE OF METALLIC COHESION

Materials cohere because their constituent atoms release energy when they combine. This energy is in turn to be associated with a change in the energy of motion of the electrons and a change in the electrostatic interaction energy of the charged constituents of the atoms, namely the positively charged nuclei and negatively charged electrons. Calculation shows that the largest source of cohesion results from the fact that the orbital valence electrons are able to take better advantage of the attractive field of the oppositely charged nuclei in the compound than in the free atom. In other words the electrons find themselves in more strongly attractive fields, on the average, in the bound state. Thus when two atoms approach one another the electrostatic fields of the two nuclei add together and increase the attractive potential field of the electrons. The magnitude of this increase and its geometrical configuration depend upon the nature of the atoms which are brought together and determine the manner in which the electrons alter their distribution.

In the case of the ideal metals, careful examination shows that the attractive field increases more or less uniformly throughout the atoms as they are brought together. As a result, the atoms arrange themselves so that each has as many neighbors as possible in order to obtain the maximum possible attractive field; hence the preference for close-packed structures. When the atoms are in contact, that is have the normal spacing, the valence electrons are able to move from atom to atom and can no longer be said to belong to particular atoms. The fields are always somewhat stronger in the vicinity of the more electronegative constituents of an alloy. As a result the electrons have a slight tendency to accumulate at the positions of these atoms. As we saw in Section III, this behavior explains, at least in part, the ability of certain alloys to become ordered. It is to be noted that in salts, such as sodium chloride, the electro-negative constituents are so much more strongly so that the "heaping" of electrons at these atoms is a predominant effect and determines the structure and their properties more definitely than any other feature of the binding. It should be emphasized that the origin of cohesion is the same in metals and in salts; the differences in the properties of the two types of material depend upon the fact that the same basic principles lead to somewhat different electron distributions in the two cases.

Associated with the circumstance that the attractive field is almost uniform from atom to atom in a metal is the fact that this field does not alter greatly when one atom is displaced slightly relative to its neighbors. In other words cohesive energy is relatively insensitive to slight distortion of the lattice. This characteristic property of the atoms in metals underlies their ductility. Large forces are not needed to produce deformation, nor do large internal stresses develop as a result of distortion. In a somewhat qualitative way, it might be said that the free electron gas flows about to equalize internal stresses as they develop during distortion.

Ideal abrasive materials, which, like salts, stand in sharp contrast with metals, have the property that the attractive potential field of the nuclei is very strong on lines that join neighboring atoms and are relatively weak in other regions. As a result the electrons have a tendency to accumulate along these lines and form what the chemists call the "homopolar bond." In these materials the energy of the lattice is very sensitive to the relative positions of the atoms, so that they are hard instead of ductile. It is to be emphasized again that there are no basic differences in the principles which determine this behavior and those which determine the fact that the electrons accumulate around the electronegative atoms in salts and are more or less uniformly distributed throughout the lattice in metals.

The theory underlying cohesion has been used to calculate the heat of sublimation of some of the simpler metals in a quantitative way. The heat of sublimation represents the energy which is required to remove an atom from the lattice and hence is an excellent measure of the cohesive energy. The excellent agreement between theory and observation in these cases leaves no doubt that wave mechanics is capable of providing a satisfactory account of the quantitative features of cohesion when the equations can be solved accurately. Unfortunately accurate solutions are not possible at present for the materials having

greatest practical interest and it is necessary to be contented with relatively approximate treatments. The cohesion of any system can always be discussed in terms of the energy levels of the constituent electrons. The most stable configuration of atoms is that in which the average energy of the electrons is lowest. Jones has shown that the dependence of the properties of the ideal substitutional alloys upon the ratio of the number of valence electrons to the number of atoms can be understood on this basis. In any one of these alloys the spectrum of energy levels depends primarily upon the structure, whereas the extent to which the levels are occupied depends primarily upon the electron-atom ratio. The most stable structure is that in which the levels are filled most efficiently. Jones has found that the observed structures of the substitutional alloys actually are those in which the levels are most efficiently filled when the electron-atoms ratio have the corresponding observed values.

VII. MAGNETIC PROPERTIES

The magnetic properties of metals have long been the source of considerable scientific as well as practical interest. They may be divided into three categories depending upon the manner in which the materials are attracted or repelled from a magnetic pole. Materials which are repelled are called diamagnetic. Those which are attracted but in which permanent magnetic poles cannot be induced are called paramagnetic, whereas those which are attracted and in which permanent magnetic poles can be generated by appropriate treatment are called ferromagnetic since iron and its compounds exhibit this phenomenon to a very marked degree.

The investigations of modern physics have demonstrated that diamagnetism is a universal attribute of moving electrons which results from the way in which the magnetic field alters their orbital motion. On the other hand paramagnetism and ferromagnetism originate in the spin of the electron; the spin becomes aligned in the magnetic field in such a way that the electron and hence the atoms to which it is attached is attracted toward the magnet.

In most of the simpler metals, which do not exhibit ferromagnetism or strong paramagnetism, pairs of electrons having opposite spin possess permissible wave patterns, in accordance with the exclusion principle. One of these electrons can reverse its spin only if it makes a transition to another energy state under the influence of a magnetic field. This can be done readily only by those electrons which lie very close to the unoccupied levels so that the paramagnetic tendency is relatively weak. Moreover, the reverse transition occurs when the field is removed. In cases such as this there is a strong competition between the diamagnetic and paramagnetic tendency and either may predominate depending upon the finer details of the behavior of the electrons.

In the ferromagnetic metals, on the other hand, certain electrons possessing orbits that lie nearer the nucleus than those of the outermost valence electrons occupy energy levels singly. In other words electrons having one type of spin orientation predominate over those with opposite spin. These electrons are of the same type as those which give the ferrous elements their multiple valence; they are sometimes called the *d*-shell electrons. The spins of the unbalanced electrons may be permanently aligned throughout the specimen by giving it proper treatment, whereupon it becomes a permanent magnet. Even if the specimen at hand has not been given this treatment, however, so that it is magnetically "soft," the spins of electrons in very large regions or *domains* are aligned parallel to one another. In such cases the specimen resembles a more or less random mixture of small magnets which are sufficiently large to be resolved under a microscope. The process of magnetization merely requires the alignment of the spins in neighboring domains.

The strongly paramagnetic metals have properties which lie intermediate between the ferromagnetic ones and those which are diamagnetic or weakly paramagnetic. In the normal state of these metals the electron spins are balanced over domains that are much smaller than those observed in the ferromagnetic materials so that complete magnetization is not easy to induce. On the other hand reversible spin transitions occur more readily than in the less strongly paramagnetic materials, so that the magnetism is stronger. The study of these materials shows that the electrons responsible for strong paramag-

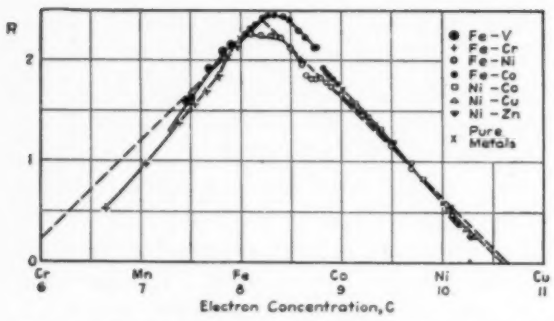


FIG. 11. The saturation magnetization of various ferromagnetic alloys as a function of the electron-atom ratio (after Shockley and Slater). It is to be noted that the various materials possess curves which nearly coincide when plotted in this manner. This illustrates the fact that the electron concentration determines the magnetic properties of this closely related family of alloys.

netism are the same as those responsible for ferromagnetism, that is, they are the electrons in orbits which move closer to the nuclei than the ordinary valence electrons and hence are less affected by chemical binding. They may be discussed in terms of a spectrum of energy levels which though discrete are much more dense than the levels of the conduction electrons. In all respects other than density, these levels resemble those of the valence electrons. Thus the electrons near the top of the filled region are responsible for the paramagnetism. Since the density of levels is high in this region more electrons can reverse their spin in a given magnetic field than is possible for the ordinary conduction electrons.

It is to be noted that if the top of the occupied region of levels is separated by a gap from the unoccupied ones, and if all of the occupied levels are completely filled in accordance with the exclusion principle, it will be difficult to reverse the spins of any electron since a relatively large amount of energy must be supplied to change the wave pattern of the electron which reverses its spin. In this case the paramagnetism vanishes and only the diamagnetic effect remains. It follows that the ideal electrical insulator should be diamagnetic, in accordance with observation.

The factors which determine whether a metal should be ferromagnetic or strongly paramagnetic are only partly understood at the present time. The difficulties are believed to arise from complexities accompanying the application of the theory rather than in its basic correctness. It is known that the ferromagnetic state is stable only if the interatomic spacing exceeds a certain value that depends upon the diameter of the orbits in the free atom; radimentary calculations of the stability have been made. It has also been shown that ferromagnetism should disappear at sufficiently high temperatures, as if found to be the case.

Figure 11 shows the saturation magnetization of various ferromagnetic alloys as a function of the electron-atom ratio. The saturation magnetization determines the optimum strength of a permanent bar magnet that is made of the material. It will be observed that the curves for the different alloys, which are of the substitutional type, superpose almost identically upon one another when plotted in this way. This result shows that the electron-atom ratio is as basic a quantity for determining the magnetization of substitutional alloys as for determining other properties.

The origin of the domains of ferromagnetic materials has been the subject of much investigation during the last fifteen years and many of the most important facts concerning this aspect of the theory of magnetism are now understood.

VIII. THE MIGRATION OF ATOMS IN METALS

Many important processes in metallurgy depend upon the fact that the atoms in metals are able to diffuse or migrate through the solid structure. For example, the hardening of steel or duraluminum by heat treatment depends upon the fact that the rate of migration of atoms can be controlled by varying the temperature. Similarly the process of case hardening of iron by introducing a layer of carbon or nitrogen into a

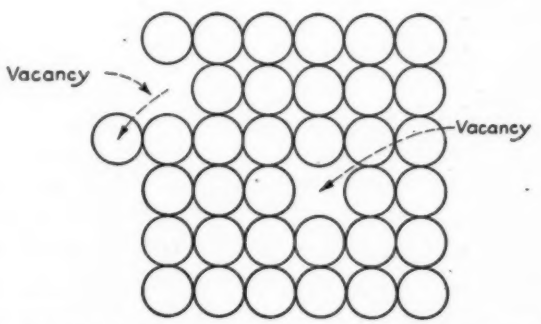


FIG. 12. Lattice vacancies. The vacancies are formed by migration of atoms in the surface planes of the crystal to the surface. It is believed that diffusion in substitutional alloys takes place with the aid of migration of vacancies.

surface layer depends upon the fact that carbon atoms will diffuse into iron at elevated temperatures as a result of the acquisition of thermal energy.

The contribution of modern physics to the understanding of the diffusion process consists in the clarification of the mechanism by which atoms jump from one position in the lattice to another. A large number of independent sources contribute to this understanding. For example, the results of x-ray analysis and wave mechanics provide methods of determining the energy of distortion that accompanies each step of migration. The theory of the thermal vibration of the atoms in solids is used to determine the way in which the migrating atoms receive the energy of thermal agitation which permits them to move.

It is generally presumed that the interstitial constituents of interstitial alloys migrate from one interstitial position to another. No evidence has accumulated to contradict this simple interpretation of the process. On the other hand, the process of diffusion in substitutional alloys has been the object of much discussion and the matter is still unsettled. The simplest assumption is that pairs of neighboring atoms in the lattice simply exchange places, forcing the surrounding atoms out of position during the interchange process. An alternative picture of the process assumes that the observed migration depends upon the presence of vacant lattice sites (Fig. 12) which occur at elevated temperatures as the result of thermal agitation. Each time such a vacant site shifts its position, an atom of the lattice is able to move. The arguments supporting the second, more complex, process are based on the belief that less distortional energy is involved in the process of moving a vacancy than in interchanging atoms in the normal lattice. In fact calculations based on application of wave mechanics have lent support to this belief in the case of copper; however, these calculations are subject to uncertainty because of unavoidable approximations which must be made in solving the fundamental equations.

In recent times a rather specialized method has been developed by Snoek[1] and Zener[2] which

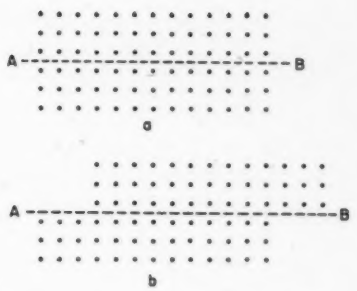


FIG. 13. Schematic representation of the end result of plastic flow in a crystal. One portion of the lattice has moved past another along a definite crystallographic place. (a) shows the undeformed crystal. (b) shows crystal after deformation, slip having occurred along the plane A–B.

may provide further insight into this problem, at least in specialized cases. This method is based on the fact that the atoms in certain alloys are induced to change their positions when the metal is placed under stress in order to achieve states of lower energy. Suitable theory shows that a sound wave passing through the metal will be relatively strongly absorbed if the frequency of oscillation coincides with the frequency with which the atoms jump from one position to another. Thus the study of the damping of sound makes it possible to determine characteristic properties of the migration of atoms. Snoek used this method at first to determine the frequency with which the carbon atoms migrate in iron-carbon alloys and Zener has extended the method for the case of the substitutional alloy alpha-brass.

IX. THE MECHANICAL PROPERTIES OF METALS

The practical interest of metallurgists is focused more upon the mechanical properties of metals than upon any other attributes. This is a direct consequence of the fact that metals possess a toughness which is unparallel among the other types of solids. It is true that some of the nonmetals, such as diamond, are much harder, but none are able to withstand under varying conditions the general type of stresses which metals can. One of the greatest contributions which physics can hope to make to the field of metallurgy is to provide a better understanding of the origin of the mechanical properties of metals and thereby provide a better basis upon which to provide control than is now available.

[1] J. Snoek, Physica **6**, 591 (1939); **8**, 711 (1941); **9**, 862 (1942).
[2] C. Zener, Trans. A.I.M.E. **152**, 122 (1943); Phys. Rev. **71**, 34 (1947).

A simple perusal of the techniques developed thus far by physicists for approaching this general problem shows that progress is certain to be slow and a revolution will not take place without continued investigation. The conventional metallurgical approach, based upon painstaking trial and error work, has yielded a large number of practical gains. In contrast, the methods employed by the physicist are relatively indirect, at least at present. Although the principles gained from modern physics make it possible to understand in a straightforward way why metals are more ductile than other materials, the principles are not yet refined to a point at which the relatively subtile differences between different metals, which so often are of major importance from the practical standpoint, can be understood clearly. There are two reasons for this: First, the structures which determine the secondary differences in the mechanical properties of metals occur on a scale of size intermediate between atomic dimensions and those visible under the microscope. For this reason they are difficult to observe in any direct way by the means available at present; in the second place wave mechanics, which is basically capable of providing answers to many of the important problems cannot be applied with sufficient accuracy at the present time. The situation may be remedied by the development of computational techniques which will make it possible to determine the interesting mechanical properties of at least the simplest metals directly from wave mechanics with recourse only to those types of experimental information which are available at the present time. This appears to be highly unlikely. It is much more probable that a combination of the present method of applying the theory and the development of experimental techniques for probing the properties of metals further will gradually provide us with a deeper understanding.

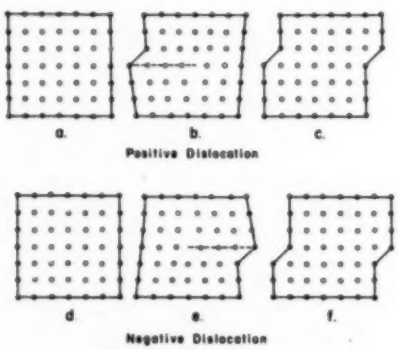


FIG. 14. Mechanism of slip with the use of dislocations (after Taylor). In the sequence (a), (b), and (c) a dislocation enters the crystal on the left hand side of the specimen and produces slip by moving from left to right. In the sequence (d), (e), and (f) the same end result is obtained by motion of a dislocation of opposite sign from right to left. In actual cases dislocations are probably produced in pairs of opposite sign in the interior of a specimen or very near to flaws where the applied stress is magnified.

The most pertinent facts which are known concerning the mechanical properties of metals at the present time are as follows:

(1) Metals, like most other solids, possess an inherent strength that is many times larger than the strength commonly observed. The theoretical ultimate strength should be of the order of one-tenth of the elastic modulus. This corresponds to a value of about 3 million pounds per square inch for iron. It is believed that the difference between the observed and theoretical strengths of materials is a result of the presence of fissures and flaws which produce regions in which the applied strength is greatly magnified. The theoretical strength would be approached if the flaws could be avoided or removed by suitable processing of the metal. In agreement with this point of view is the fact that freshly drawn fibers of glass which presumably possess few flaws are very strong and the fact that the strength of metals increase as they are cold worked under conditions of compression which would permit flaws to be diminished in size and welded together. For example, Beams has found that small steel balls of the highest quality possess bursting strengths in the amounts of 500,000 pounds per square inch when centrifuged at higher speed of rotation.

(2) A combination of x-ray and mechanical analysis shows that planes of the lattice slide past one another when metals flow ductily (Fig. 13). Such flow requires the existence of stresses which act in the plane of flow. Generally speaking, displacement occurs most readily along the planes having the highest atomic density. Theoretical analysis shows that the observed stresses for which plastic flow is observed are too small to be explained on the assumption that neighboring atomic planes move past one another as rigid units, the displacement occuring over the entire area of the planes at once. Instead, it must be

concluded that a wave of relative displacement passes from one side of the area to the other. The simplest picture of the displacement process is that shown in Fig. 14 which makes use of a form of lattice imperfection which is called a dislocation (Fig. 15). The disorder shown in the single crystallographic plane in the figure extends through many planes parallel to the one illustrated. Thus the loci of the centers of the disorder in each of these planes form a line which is called a *dislocation line*. These lines need not be straight and in fact probably form closed curves. Once formed these imperfections can be made to move across the crystallographic planes by application of relatively low stresses; the end result of such motion is exactly the effect observed during plastic flow, namely the relative displacement of crystallographic planes. It is not yet known whether the dislocations responsible for plastic flow are always present in the crystal in the form of unavoidable imperfections or whether they are generated at regions of high stress during flow.

(3) The relatively great ductility of metals is intimately associated with the nature of the bonding forces in these solids. As was mentioned in Section VI, the energy of the atoms in metals is less sensitive to minor displacements than is the case for non-metals. Hence the rupture strength is not exceeded when the metal is stressed in such a way that dislocations are generated and set in motion.

The theory of dislocations has been developed by a number of investigators and it appears that many of phenomena observed in connection with plastic flow can be explained with their use. For example, it is possible to explain the hardening that occurs and the fact that a fraction of the energy expended during the process is stored. Similarly it is possible to explain the fact that alloys are less ductile than pure metals and that

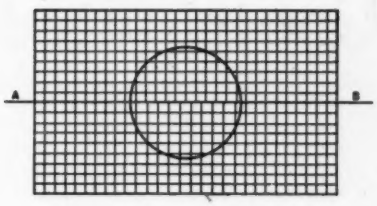


FIG. 15. A schematic illustration of a dislocation (after Orowan). It will be noted that there is one more vertical line of atoms above the slip-plane $A-B$ than below it so that the arrangement of cells above and below the line $A-B$ resembles the arrangement of scale lines in a Vernier scale. The center of the dislocation will thus move a unit lattice distance when the atoms move a small fraction of this distance. It is to be emphasized that the dislocation pattern shown extends through many lattice planes above and below that shown.

hardening occurs during the precipitation of a phase in a binary system such as duraluminum or steel. In spite of these successes the theory undoubtedly must be expanded in the future before our understanding can be said to be complete. This is one of the most important problems of the physics of metals.

BIBLIOGRAPHY

1. C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc. New York, 1943).
2. J. W. Beams, paper presented to Southeastern Section of American Physical Society, April, 1947.
3. W. Hume-Rothery, *Structure of Metals and Alloys* (New York Chemical Publishing Company, New York, 1939).
4. R. P. Johnson and F. Seitz, J. of App. Phys. **8**, 84, 186, 246 (1937).
5. N. F. Mott and H. Jones, *The Structure of Metals and Alloys* (Oxford Press, Cambridge, 1936).
6. F. Seitz, *The Physics of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940).
7. W. Shockley, Bell Sys. Tech. J. **18**, 645 (1939).
8. J. C. Slater, Rev. Mod. Phys. **6**, 209 (1934).
9. C. Zener, *Elasticity and Anelasticity of Metals* (Chicago University Press, Chicago, 1948).

Contributed Original Research

Velocity of the Anode Spark in Copper Sulfate Solutions Under Application of Impulsive Potential

HUGH F. HENRY
Department of Physics, University of Georgia, Athens, Georgia

(Received February 27, 1948)

The velocity of the luminous-anode spark produced in $CuSO_4$ solutions by application of impulsive potential has been measured as a function of the interelectrode gap, input voltage, and solution normality. Determinations were made for input potentials of 15–30 kv, interelectrode gaps of 15–80 mm, and solution concentrations of 0.0125–0.0625N.

The spark velocity was found to depend upon the input potential and the interelectrode gap by the empirical equation:

$$V=311.1(E_0-11.7)/G,$$

where V is the spark velocity in meters/sec., E_0 is the input voltage in kv, and G is the interelectrode gap in cm. Qualitatively, it was observed that as the solution normality increased, the maximum breakdown distance decreased and the spark behaved more erratically.

The maximum length which the spark streamers reach may be stated empirically by the equation:

$$L_m=2.33\times10^{-3}(E_0+2.89)^2/(N+0.545)(G-0.43)^{\frac{1}{2}},$$

where L_m is the maximum streamer length in cm, E_0 is the input potential in kv, G is the interelectrode gap in cm, and N is the normality of the solution.

The computed potential at which the spark ceases its progress through the liquid, E_m, was found to be proportional to the input potential but independent of the interelectrode gap and the solution concentration. The ratio E_m/E_0 was 0.646. The time lag in initiation of the spark was also found to depend upon the input voltage but to be independent of the interelectrode gap and the solution concentration.

INTRODUCTION

IN the general field of electrical discharge phenomena, most of the projects involving liquids have been devoted to a study of dielectric materials of low conductivity, principally in an effort to determine the cause of breakdown and consequent dielectric failure of the liquids under test.

The recent theories of Seeger and Teller[1a] and of Frohlich[1b] consider that breakdown is an ionic process in which the dielectric contains a certain number of electrons at least partially free of molecular binding. These "free" electrons move under the influence of an applied electric field and excite vibrations of the positive and negative ionic constituents of the material under test. For field strengths below the breakdown value, the energy received from the field by the electrons is completely absorbed by these ionic oscillators with the result that their velocity is held to a low value. Stronger fields, it is assumed, can accelerate the electrons until they acquire energy sufficient for ionization, which in this case implies the release of an electron from the negative ion. From their experiments with CCl_4, Attwood and Bixby[2] came to the conclusion that these theories may be applicable to a wide range of dielectric materials, provided their constitution is ionic. It is interesting that although their results are undoubtedly valid, the material which they used, CCl_4, is considered chemically to have covalent bonds and to be non-ionic in character.[3]

Snoddy and Beams[4] found that electrical spark discharges could be initiated in copper sulfate

[1] (a) R. J. Seeger and E. Teller, Phys. Rev. **54**, 515 (1938); (b) H. Frolich, Proc. Roy. Soc. **160**, 230 (1937).

[2] S. S. Attwood and W. H. Bixby, J. Frank. Inst. **235**, 259 (1943).

[3] For example, see S. Sugden, *The Parachor and Valency* (George Routledge & Sons, Ltd., London, 1930), p. 90 ff. or J. F. Norris and R. C. Young, *A Textbook of Inorganic Chemistry for Colleges* (McGraw-Hill Book Company, Inc., New York, 1938), second edition, p. 474 ff.

[4] L. B. Snoddy and J. W. Beams, Phys. Rev. **55**, 879 (1939).

solutions. The author found that the same was true for solutions of other strong electrolytes.[5] Accordingly, an investigation of some of the spark properties of strong electrolytes has been initiated. This paper is a report of one phase of that investigation, namely, a study of the velocity of the luminous spark produced at the anode in copper sulfate solutions under the application of impulsive potential.

APPARATUS AND METHOD

The discharge cell consists of a wooden frame with clear celluloid sides approximately 0.58 cm thick. It was made watertight with a rubber gasket and aquarium cement at the celluloid-wood joints. The wood itself was made leakproof by treatment with paraffin, and the interior wooden parts of the cell were coated with paraffin. The inside dimensions of the cell were 17.0 cm deep, 17.0 cm wide, and 9.3 cm thick. The electrodes were copper rods, 0.64 cm in diameter, with rounded conical points. They were thrust through the wooden ends of the cell so that the discharge path between their tips was 6.3 cm from the cell bottom and 4.6 cm from each of the celluloid sides. When 2 liters of solution were placed in the cell, the discharge path between the electrodes was 5.8 cm below the liquid surface. The dimensions of the cell were such that spark photographs taken through the side of the cell and from above showed no effects of the sides or free surface. Thus all sparks were three-dimensional in character.

It was found impracticable to use an all-glass discharge cell, as complete breakdown of the liquid produces a powerful compressional wave which smashes such a cell. For the same reason, plate glass sides for the cell were discarded; however, test experiments indicate that the discharge properties being studied were the same for all three cell types. Similarly shaped electrodes of brass, aluminum, iron, and carbon were used with no detectable difference in spark velocity characteristics being noted.

Figure 1 shows photographs of the sparks produced in liquids under different conditions. All photographs of the sparks were made by a camera with an f:4 lens and using 616 film. Although Eastman Super-XX and Verichrome and Afga Superpan Press and Plenachrome films were used in the spark photography, no detectable differences in the spark appearance caused by the film used were noted.

The discharge was produced by the impulse generator shown in Fig. 2. The parallel circuit consisting of R_1, R_2, C_1, S_1, and S_2 was designed to remove the potential applied across the discharge cell after a calculated time interval had

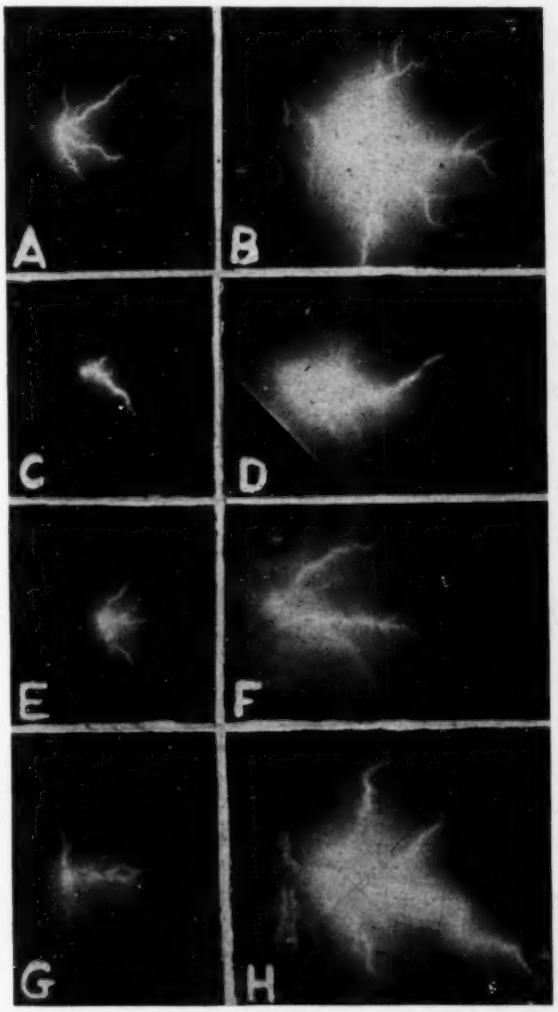


FIG. 1. Actual photographs of the sparks made under the following conditions:

Photo	E_0(kv)	$C(N)$	G (cm)	Time (μs)
A	20	0.0250	4.0	8.0
B	20	0.0625	4.0	8.0
C	15	0.0250	4.0	8.0
D	25	0.0250	4.0	8.0
E	20	0.0250	6.0	8.0
F	20	0.0250	4.0	Inf.
G	20	0.0250	4.0	12.0
H	20	0.0625	4.0	Inf.

[5] Unpublished work done by the author at the University of Virginia.

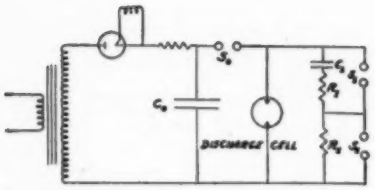


FIG. 2. Circuit of impulse generator. $C_0=0.25$ mf; $C_1=3.00\times10^{-4}$ mf; $R_1=8.3\times10^3$ ohms; $R_2=2.50\times10^4$ ohms. Other circuit elements have conventional values. In the text, the resistance of the solution in the discharge cell is called R_g.

passed. It will be identified as the cut-off circuit. R_g is the resistance of the solution in the discharge cell which actually varied in the experiments from 50 to 300 ohms, depending upon the concentration of the solution and the interelectrode gap distance. Since R_1 and R_2 are much higher than R_g in all cases, the voltage (E_g) applied across R_g, and consequently the cut-off circuit, could be considered an exponential voltage produced by the discharge of C_0 through R_g alone and given by the relation, $E_g=E_0\exp(-kt)$, where k is $1/R_gC_0$ and E_0 is the potential to which C_0 is charged and is thus the initial voltage applied to the discharge cell.

Since C_1 is charged through R_1 and R_2 by that exponentially decreasing voltage, the potential across spark gap S_1 may be mathematically determined to be given by the equation:

$$E_1=[E_0/(1-RC_1k)][(1-R_1C_1k)\exp(-kt) -(1-R_1/R)\exp(-t/RC_1)],$$

where R is R_1+R_2. For voltages above 20 kv, C_0 was replaced by a two-stage Marx circuit. Fig. 3 shows two graphs of E_1 for different values of R_g. Also shown are the corresponding curves of E_g.

From an analysis of the equation for E_1 and the curves shown in Fig. 3, it may be seen that the voltage across S_1 will reach a maximum which depends upon the value of R_g. If S_1 is set for a potential above that maximum, it will not break down. The value of R_g to be used in any set of experiments was determined experimentally by finding the voltage at which the cut-off circuit failed to operate and from that value finding graphically the corresponding value of R_g.

In practice, S_1 is adjusted so that it will break down after a predetermined time interval. At its breakdown, S_2 is highly overvolted, and so the triple spark gap circuit, S_0, S_1, and S_2 discharges C_0 very rapidly and thus removes the voltage from R_g. The three gaps, using 2.5-cm brass balls, were very compactly placed to eliminate as much inductance in their circuit as possible, and were illuminated with ultraviolet light from a quartz mercury arc. Oscillographic investigation of the breakdown showed that the potential was removed from the electrodes in approximately 10^{-7} sec., and that the maximum surge of the highly damped oscillation produced at the discharge of the condenser through the spark gap circuit was less than 15 percent of the input potential.

Experimentally, a series of photographs of the sparks produced for different time intervals of application of potential was taken for a given set of conditions. It was assumed that the extension of the spark into the liquid ceased when the voltage was removed from the discharge cell. The maximum lengths of those sparks were measured and plotted as a function of the time for which the voltage was applied; the slope of the resultant curve was considered to be the spark velocity. Figure 4 shows a typical set of data thus plotted. Since all of the curves produced by this method, with a few exceptions at 30 kv, could be best fitted by a straight line, it was assumed that in all cases the velocity of the spark was constant after an apparent initial period of high velocity indicated by AB in the graph. The spark velocity was found by determining the slope of that line (BC in the diagram) by the method of least squares.

The time lag in the starting of the spark (point A in Fig. 4) was determined in each case by find-

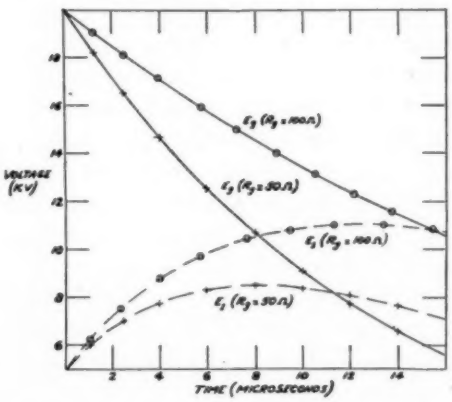


FIG. 3. Values of E_1 (potential across spark gap S_1) and E_g (potential across discharge cell) for different values of R_g.

ing the shortest time for which the cut-off circuit could be set for a spark to be observed visually in a darkened room. If the potential were removed 5×10^{-7} sec. earlier, the spark was not observed.

It was found that a spark could start at either electrode, with the time lag in its initiation depending upon the radius of curvature (or bluntness) of the electrode tip. In order to obtain sparks which started in a sufficiently short time to be measured by the method described, it was necessary to use an electrode whose tip had a very small radius of curvature. The initial field at the tip of an electrode satisfying the conditions named was of the order of 10^5–10^6 volts/cm for an applied potential of 20 kv. This value was computed by the method of Eyring, MacKeown, and Millikan[6] from photographs of the actual electrode tip. The main disadvantage of such a sharply pointed tip was the fact that a single breakdown pitted and blunted it to such an extent that the time lag in starting of the spark was changed. The streamers produced after such a breakdown would thus not lie on the same curve as those produced before.

It was found that if a gaseous bubble were placed upon the electrode tip, either by mechanical means or by a pre-discharge electrolysis with 110-volt direct current applied across the electrodes for about $\frac{1}{4}$ sec., the time lag in starting of the spark was the same as that for a sharply pointed electrode regardless of the actual shape of the electrode used.[5,7] The velocity characteristics of the spark produced are the same in the two cases. Consequently, for most of the experiments the spark was initiated by placing a bubble upon the electrode tip, although sharply pointed electrodes were used in certain cases where there was no danger of breakdown.

RESULTS

Spark velocities were measured for an input capacitance of 0.25 mf and as functions of the input potential, solution concentration, and interelectrode gap. The ranges studied are given below:

Input potential (kv)	Solution concentration (normality)	Interelectrode gap (mm)
15	0.0250–0.0375	15–40
20	0.0125–0.0625	15–60
25	0.0125–0.0625	15–80
30	0.0125–0.0313	15–80

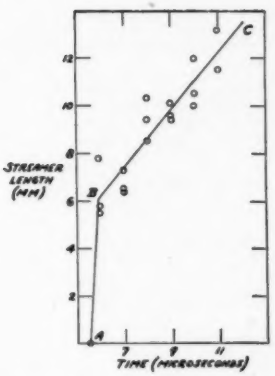


FIG. 4. Typical set of experimental results.

It was found that for values of the concentration outside the range given for 15 kv, and above the maximum for 30 kv, the spark did not start in the time interval for which the cut-off circuit was effective. The same was true for a few short interelectrode gaps at high concentrations within the above ranges.

For interelectrode gaps of 10 mm and for a few 15-mm gaps, the spark apparently bridges the gap in less than 5×10^{-7} sec.; this indicates a spark velocity of the order of 2×10^4 meters/sec.

Table I shows a typical set of data determined by the method outlined above. The column headed number points gives the actual number of streamers whose lengths were measured in finding the spark velocity. The column headed

TABLE I. Input capacitance -2.5×10^{-7} farad; input voltage -25-kv; solution concentration $-0.0250N$.

Elec. gap (mm)	No. points	Spark starts (μs)	Max. time (μs)	Minimum spk. len. (cm)	Maximum spk. len. (cm)	Velocity (meters/sec.)	Prob. error (m/sec.)	E_m (kv)
15	8	5.5	7.0	1.13	1.50(*B*)	2780.0	221.7	
20	8	6.0	9.0	1.07	2.00(*B*)	3235.0	264.0	
25	10	6.0	10.0	0.95	2.50(*B*)	2075.0	534.1	
30	15	6.0	10.0	0.83	1.87	700.0	132.7	12.8
40	12	6.0	10.0	0.72	1.60	976.1	30.3	16.8
50	10	6.0	12.0	0.63	1.41	732.1	99.4	16.8
60	15	6.0	16.0	0.61	1.33	537.7	60.0	16.0
70	16	6.0	16.0	0.57	1.32	382.9	80.7	15.0
80	17	6.0	18.0	0.42	1.10	427.2	52.9	16.4

[6] Eyring, MacKeown, and Millikan, Phys. Rev. **31**, 900 (1928).

[7] Snoddy, Henry, and Beams, Phys. Rev. **57**, 350 (1940).

TABLE II.

Input volt (kv)	No. terms	E_m/E_0	Probable error of mean
15	12	0.611	2.93%
20	28	0.690	1.29%
25	35	0.648	1.42%
30	13	0.573	5.24%
All	88	0.646	1.05%

spark starts is the time lag in the appearance of the spark (point A in Fig. 4). It was found to depend slightly upon the applied potential but not upon either the solution concentration or the interelectrode gap. Averages are as follows:

Input voltage (kv)	Time lag (μs)
15	7.9
20	7.0
25	5.9
30	9.0

The column headed maximum time shows the maximum time for which the progress of the spark was measured; in most cases, it was the same as the longest time for which the cut-off circuit was operable with the given circuit constants and solution resistance. The time interval for which the spark velocity was measured is obviously the difference between the values in the two columns headed spark starts and maximum time.

The column headed minimum spark length states the average length of the spark when first observed (point B in Fig. 4), while the column headed maximum spark length gives the average length of the spark observed when the cut-off circuit was disconnected. The letter B in this column shows that the spark bridges the interelectrode gap and breakdown occurs.

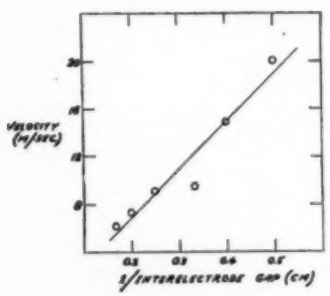


FIG. 5. Average spark velocities as function of inverse of interelectrode gap.

TABLE III.

Input volt (kv)	No. terms	*R	Probable error of mean
15	15	270.8	8.68%
20	36	344.5	5.92%
25	46	298.3	5.23%
All	97	311.1	3.62%

* $R=V(G)/(E_0-11.7)$.

The columns headed velocity and probable error are self-explanatory. Since the number of points used in determining each velocity is small, the values in the latter column are really the statistical confidence intervals for 50 percent. The probable errors shown in this table are about average for all of the data.

The column headed E_m gives the computed values of the voltage existing across the interelectrode gap when the spark luminosity reaches its greatest length. In finding it, the measured velocity was assumed constant and the time, t_m, for the spark to reach its greatest length was determined from the average maximum length of the spark and its velocity. Since $E_m=E_0\exp(-t_m/R_gC_0)$ and R_g and C_0 are known, E_m may be computed.

Determinations of E_m under all conditions studied indicate that it is independent of the interelectrode gap and the solution concentration but is proportional to the input potential E_0. A summary of the values of E_m/E_0 is given in Table II, where it is shown that the average value of the ratio is 0.646 ± 1.05 percent.

The small error associated with the average value of E_m for each potential indicates that the original assumption that the spark velocity is constant during its existence is probably valid. Also, with C_0 constant, the fact that E_m/E_0 is also constant indicates that t_m is proportional to R_g.

To find the effect upon the streamer velocity of one of the variables studied (interelectrode gap, solution concentration, or input potential), all of the velocities for one value of that variable were averaged without consideration of the other variables. The velocity averages thus obtained were plotted as a function of the variable where the function was so chosen as to form a straight line graph. This method eliminated the effect of statistical fluctuations in the velocity determinations.

Figure 5, where the velocity is plotted as a function of the inverse of the interelectrode gap, shows the effect of changing that gap. Voltage and solution concentration values were used only if they included velocity measurements in the complete range of 20 mm to 60 mm, which was the range most suitable for determining this effect. Each point on the graph is thus the average of 13 individual velocity determinations which include all concentrations at 30 kv and all but those for a $0.0625N$ concentration at 20 and 25 kv.

The equation of the straight line shown is of the form $V=k/(G-0.007)$, where V is the velocity in m/sec., G is the interelectrode gap in cm, and k is a constant. Since the factor 0.007 is negligible in comparison with the values of G used and well within the experimental errors, it may be neglected and the conclusion thus reached that the spark velocity is inversely proportional to the interelectrode gap.

In a similar manner, the dependence of the velocity upon the input potential was found by using the results obtained for interelectrode gaps of 40 mm and less and for solution concentrations of $0.0375N$ and less for each potential. The result is plotted in Fig. 6, where each point is the average of 20 individual velocity measurements. The equation of the straight line drawn between 15 and 25 kv is of the form $V=K(E_0-11.7)$, where V is the velocity in m/sec., E_0 is the input potential in kv (not greater than 25 kv), and K is a constant. It is not obvious why the point at 30 kv indicates a smaller velocity than that for 25 kv. Qualitative observations seem to indicate that at 30 kv the spark discharge in its initial stage (which is the region of measurement by the method described) may be at least partially nonluminous. It should be noted that the luminous spark starts too late to be measured by the method used at solution concentrations greater than $0.0313N$ for 30 kv.

The effect of the solution concentration can be stated only qualitatively. For interelectrode gaps less than 40 mm, the velocity increases with an increase of solution normality. For gaps greater than 40 mm, the velocity reaches a maximum at about $0.0375N$ and then decreases as the normality is increased. Both effects are relatively small.

TABLE IV.

Input volt (kv)	No. terms	$*P\times10^3$	Probable error of mean
15	26	2.450	2.48%
20	92	2.445	1.19%
25	112	2.030	1.14%
30	85	2.558	1.13%
All	315	2.328	0.70%

$* P=L_m(G-0.43)^{\frac{1}{2}}(N+0.545)/(E_0+2.89)^2$.

At higher concentrations the time lag in spark initiation increases as the interelectrode gap is decreased. The probable error in determination of the spark velocity is much larger for the high concentrations than for the lower ones, indicating that the spark behaves more erratically as the solution concentration increases.

The above analysis indicates that the velocity of the spark in $CuSO_4$ solutions of the normality studied may be expressed by the empirical equation:

$$V=R(E_0-11.7)/G,$$

where V is the spark velocity in m/sec., E_0 is the input potential in kv (not greater than 25 kv), G is the interelectrode gap in cm, and R is a constant.

R was determined from each experimental value of the velocity between 15 kv and 25 kv, and its average was found to be 311.1 ± 3.62 percent. A summary of its values is given in Table III. The comparatively large error associated with it is probably due to the unknown effects of the solution normality.

In a similar manner, the dependence of the maximum streamer length upon the three variables was determined and the following empirical equation found:

$$L_m=P(E_0+2.89)^2/(N+0.545)(G-0.43)^{\frac{1}{2}},$$

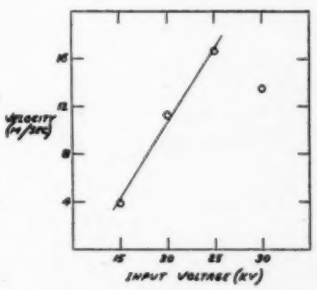


FIG. 6. Average spark velocities as function of applied potential.

where L_m is the maximum streamer length in cm, E_0 is the input potential in kv, N is the normality of the solution, G is the interelectrode gap in cm, and P is a constant. P has an average value of $2.328\times10^{-3}\pm0.70$ percent. The results are summarized in Table IV.

It should be noted further that the maximum breakdown distance is proportional to the input potential and increases as the solution normality is decreased below $0.0250N$.

DISCUSSION OF ERRORS

The principal source of experimental error is probably the fact that the measured sparks are two-dimensional projections of three-dimensional figures. The photographs (Fig. 1) show that the streamers branch, and in some cases the angle between the direction of the longest spark streamer and the direct path across the gap is comparatively large. From observations of the sparks, it is concluded that this deviation is least in cases where the cut-off circuit is disconnected and the spark reaches its maximum length. The probable errors of individual terms in measurements involving the maximum streamer lengths indicate that the *minimum error* in the determination of the spark velocity caused by this factor is 10 percent.

An oscillograph was used to check the timing of the cut-off circuit under various experimental conditions and for the complete time range over which the spark's progress could be measured. The greatest probable error of a single determination of the time was found to be 6.2 percent, which was the probable error at the maximum time for which the cut-off circuit was operable. Since the probable error was less for the shorter time intervals, the over-all probable error for a single determination of the time by the method used was found to be 4.1 percent.

Personal opinion in estimating the actual position of the streamer tip and in measuring the streamer lengths could have caused an error in the value of the streamer velocity of about 7 percent. This figure was established by comparing results obtained in the measurement by R. G. Henry, E. C. Long, and the author of test films picked at random. In no case was a discrepancy in the values of the velocities found.

Since the errors resulting from the causes named above are partially compensating, it is estimated that the maximum experimental error is not greater than 15–20 percent.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the suggestions and help of the author's colleagues, E. H. Dixon, L. L. Hendren, and R. G. Henry, in the planning of the project and in evaluating the results obtained. T. A. Bancroft of the Mathematics Department aided very much in the statistical interpretation of the data obtained. Thanks are also due S. F. Thomas, who designed and constructed several special pieces of apparatus for use in the experiments.

The author wishes to express his appreciation to the members of the Advisory Faculty Council of the University Center in Georgia who made available, through a grant-in-aid, part of the funds necessary for the project.

Bidwell's Intercept Relation and the Thermal Conductivity of Liquid Metals

R. W. Powell

Physics Division, The National Physical Laboratory, Teddington, Middlesex, England

(Received November 24, 1947)

A note pointing out that the claim made by Bidwell and Hogan [Journal of Applied Physics **18**, 776 (1947)], that the relation $k/\rho C = K(1/T) + K^1$, where k is the thermal conductivity, ρ the density, C the specific heat, T the absolute temperature and K and K^1 are constants, holds for aluminium, tin, lead and zinc with the intercept, K^1, the same for the solid and liquid phase, is not so well supported by experimental evidence as the reader is led to believe.

The latest paper on aluminium makes no reference to Konno's appreciably different results for the thermal conductivity of liquid aluminium. Furthermore, adjustment appears necessary in the plotting of Bidwell and Hogan's two points for the liquid phase and this brings the intercept for the liquid phase about 30 percent above that for the solid phase.

Of the other three metals cited by Bidwell and Hogan as having been previously shown to conform to the intercept relationship, it is revealed that zinc does not on the experimental evidence of both Bidwell and Konno, lead does on the evidence of Bidwell but not on that of Konno, whilst tin does on the evidence of Konno only. Further experimental investigation is clearly necessary before any definite relationship such as that suggested can be accepted.

THE paper on "Thermal conductivity of aluminium; solid and liquid states" by C. C. Bidwell and C. L. Hogan published in the August number of the Journal of Applied Physics appeared shortly after I had been examining Bidwell's earlier contributions to determine whether his suggested intercept relation might be useful as a means whereby additional information could be obtained regarding the thermal conductivities of liquid metals. Some rather surprising inconsistencies were observed, and, as others appear in this most recent paper I feel that the matter calls for comment.

In the introduction to the above paper the statement is made that no data for the thermal conductivity of aluminium in the liquid state had been reported previously. In point of fact, Konno has published values of 0.223 and 0.214 for the thermal conductivity of liquid aluminium at temperatures of 675° and 800°C respectively, and it is difficult to see how these values could have been overlooked as Bidwell has made frequent references to Konno's work on liquid metals, and Bidwell and Hogan quote Konno's values for the solid phase. Konno's data for both solid and liquid phases were originally published in the same table and figure. Konno's values for liquid aluminium are considerably greater than Bidwell and Hogan's and decrease with temperature instead of increasing.

The value of Bidwell and Hogan's work would have been increased had more than two experimental points been given for the liquid state. Furthermore it is regrettable that a critical examination of Bidwell and Hogan's plot of $k/\rho C$ against $1/T$, where k is the thermal conductivity, ρ the density, C the atomic heat and T the absolute temperature, leads to doubts regarding the location of these two points. Taking values of C and ρ from "The Aluminium Industry," as the authors claim to have done, gives for the coordinates of the 740°C point

$$k/\rho C = 0.143/2.36(0.2543 \times 26.97) = 0.00882,$$

$1/T = 1/1013 = 0.000987$, and for the 900°C point

$$k/\rho C = 0.180/2.316(0.2619 \times 26.97) = 0.011,$$

$1/T = 1/1173 = 0.000853$, whereas, to within the accuracy of measurement, the two points plotted have coordinates of 0.0085, 0.00108 and 0.0113, 0.00079 respectively. These adjustments bring the intercept for the liquid phase about 30 percent above that for the solid phase.

Figure 1 shows Bidwell and Hogan's data replotted in the foregoing way, and also includes a similar treatment of Konno's values. It becomes

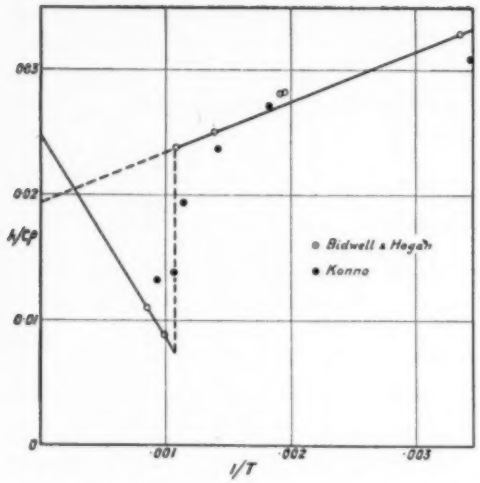


FIG. 1. $k/C\rho$ against 1/T for aluminium; replotting Bidwell and Hogan's data and including values derived from Konno's data for k. k=thermal conductivity; C=atomic heat; ρ=density; T=absolute temperature.

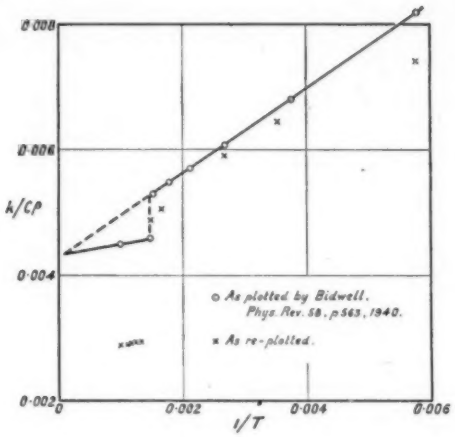


FIG. 2. $k/C\rho$ against 1/T for zinc; as originally plotted by Bidwell and replotted. k=thermal conductivity; C=atomic heat; ρ=density; T=absolute temperature.

apparent that the results for aluminium are not so consistent with the Bidwell intercept relation as the reader of this latest paper is led to believe.

Turning now to a more general consideration of the subject, the statement in the present paper, "This relation (i.e. $k/\rho C=K(1/T)+K^1$) is found to hold with the intercept, K^1, the same for the solid and liquid phase, exactly as was found for tin, lead and zinc [Bidwell, Physical Review **58**, 561 (1940)]," gives rise to the following comments regarding Bidwell's earlier treatment for these three metals.

Tin. Konno's thermal conductivity data are used to show that the intercept relation is satisfied. Bidwell has made no determinations himself.

Lead. Konno has also published values for the thermal conductivity of solid and liquid lead. In this instance the values do not satisfy the intercept relation, so a redetermination was made by Bidwell. Bidwell's values for the thermal conductivity of liquid lead are about 50 percent greater than Konno's, but as they give agreement with the intercept relation it has been concluded that the relation is satisfied by lead.

Zinc. In this instance a mistake appears to have been made which is difficult to understand. Konno and Bidwell have both determined the thermal conductivity of solid and liquid zinc and their values are in reasonable agreement. Bidwell plots his results and shows that both intercepts coincide at $k/\rho C=0.0043$. Taking a value of 7.2 for C and 6.6 for ρ leads to a value of 0.205 for the thermal conductivity of liquid zinc at $1/T=0$. As Bidwell's experimental values (read from his curve) are 0.144 at 460°C and 0.135 at 700°C there would seem to be some large discrepancy. The accompanying figure, Fig. 2, may serve to make this clear. It shows Bidwell's results for zinc as plotted in the graph in question and also as replotted using for K values read from his published curve (this occurs in Phys. Rev. **56**, 594 (1939) as well as in the 1940 paper) and for C and ρ values taken from the International Critical Tables. The replotted results show that the claim that the intercepts for the solid and liquid phases coincide is far from true. Bearing in mind that Bidwell's thermal conductivity data for zinc were in fairly good agreement with Konno's earlier determinations, it appears unlikely that the intercept relation holds good for zinc.

To sum up, of the three metals cited by Bidwell and Hogan as having been previously shown to conform to the intercept relationship, zinc does not on the evidence of both Bidwell and Konno, lead does on the evidence of Bidwell but not on that of Konno, whilst tin does on the evidence of Konno only. There is clearly need for further experimental investigations before any definite relationship such as that suggested by Bidwell can be accepted.

The Effective "Gamma" for Isentropic Expansions of Real Gases*

ARTHUR S. IBERALL
National Bureau of Standards, Washington, D. C.
(Received March 8, 1948)

In an isentropic expansion of a real gas, the relationship $pv^\gamma = C$ (γ, C constant) is not exact. A development is given for an accurate exponent $\bar{\alpha}$ (the effective "gamma"), related to the ratio of specific heats, which permits the use of the relationship $p_0v_0^{\bar{\alpha}} = p_1v_1^{\bar{\alpha}}$ for an isentropic expansion between any two pressures. The exponent may be computed from the initial conditions of the gas, the expansion ratio, equation of state data, and data on the variation of specific heat with temperature.

INTRODUCTION

IN investigating certain corrections to the elementary theory of gaseous discharge through orifices and nozzles, it was found necessary to describe accurately the states of a real gas undergoing an isentropic expansion. This involved computation of the pressure and temperature coefficients of the specific heat ratio γ from equation of state data and the temperature variation of specific heat at zero pressure. For convenience, it appeared desirable to retain the form $p_0v_0^{\bar{\alpha}} = p_1v_1^{\bar{\alpha}}$ to represent the state of a gas before and after an isentropic expansion. It was therefore necessary to relate the effective exponent $\bar{\alpha}$ to the true ratio of the specific heats γ.

ISENTROPIC EXPANSION OF A REAL GAS

The key to the solution of finding an effective exponent lay in defining and determining the properties of an auxiliary function

$$\alpha = -\frac{v}{p}\left(\frac{\partial p}{\partial v}\right)_s. \tag{1}$$

The letter p will be used to denote the total pressure, and v and s to denote the specific volume and entropy, respectively.

The relationship between γ and α can be found from the thermodynamic equation

$$\gamma\left(\frac{\partial p}{\partial v}\right)_T = \left(\frac{\partial p}{\partial v}\right)_s \tag{2}$$

to be

$$\alpha = -\frac{\gamma v}{p}\left(\frac{\partial p}{\partial v}\right)_T. \tag{3}$$

The properties of α can be determined by writing Eq. (1) in the form

$$\int_s\left(\frac{dp}{p}+\alpha\frac{dv}{v}\right)_s = 0, \tag{4}$$

and integrating as

$$\ln p/p_0 + \int_s^v\left(\frac{\alpha}{v}\right)dv = 0. \tag{5}$$

The subscript s is here used to denote that the integration is to be performed along an isentropic path.

If $\bar{\alpha}$ is defined by the relationship

$$\bar{\alpha} = \frac{\int_s^v\left(\frac{\alpha}{v}\right)dv}{\ln v/v_0}, \tag{6}$$

Eq. (5) then becomes

$$\ln p/p_0 + \bar{\alpha}\ln v/v_0 = 0, \tag{7}$$

or

$$p_0v_0^{\bar{\alpha}} = pv^{\bar{\alpha}}. \tag{8}$$

The function $\bar{\alpha}$ therefore gives the correct value of the effective exponent that may be used in describing an isentropic process of a real gas.

The function $\bar{\alpha}$ may be expanded in a Taylor series about its value for the initial state of a gas as

$$\bar{\alpha} = \bar{\alpha}^0 + \left(\frac{\partial\bar{\alpha}}{\partial v}\right)_s^0(v-v_0) + \frac{1}{2!}\left(\frac{\partial^2\bar{\alpha}}{\partial v^2}\right)_s^0(v-v_0)^2 + \cdots. \tag{9}$$

* This work was supported by the Office of Naval Research under a project on "Basic Instrumentation for Scientific Research."

The expansion given in Eq. (9) may be regarded as the Taylor expansion of a function $\bar{\alpha}$ of the two variables, v and s, for constant s. The superscript 0 is here used to indicate that the coefficients are to be evaluated at the initial condition of the gas v_0, s_0.

The coefficients $\bar{\alpha}^0$, $(\partial\bar{\alpha}/\partial v)_s^0$, etc., can be evaluated from Eq. (6) as illustrated below.

$$\bar{\alpha}^0 = \lim_{v\to v_0} \bar{\alpha}$$

$$= \lim_{v\to v_0} \frac{\int_s^v \left(\frac{\alpha}{v}\right) dv}{\ln v/v_0}$$

$$= \alpha^0. \tag{10}$$

This demonstrates that for an infinitesimal expansion of a gas (as in a sound wave), the effective exponent is given by α^0, a function of the initial state of the gas. Since, in real gases

$$\left(\frac{\partial p}{\partial v}\right)_T \neq -\frac{p}{v}, \text{ then } \alpha^0 \neq \gamma.$$

The second coefficient, $(\partial\bar{\alpha}/\partial v)_s^0$, can be evaluated as follows:

$$\left(\frac{\partial\bar{\alpha}}{\partial v}\right)_s = \left(\frac{\partial}{\partial v}\right)_s \frac{\int_{v_0}^v \left(\frac{\alpha}{v}\right) dv}{\ln v/v_0} = \frac{\frac{\alpha}{v} - \frac{\int_{v_0}^v \left(\frac{\alpha}{v}\right) dv}{v \ln v/v_0}}{\ln v/v_0}. \tag{11}$$

In the limit, Eq. (11) becomes

$$\left(\frac{\partial\bar{\alpha}}{\partial v}\right)_s^0 = \frac{1}{2}\left(\frac{\partial\alpha}{\partial v}\right)_s^0. \tag{12}$$

Since

$$\left(\frac{\partial\alpha}{\partial v}\right)_s = \left(\frac{\partial\alpha}{\partial v}\right)_T + \left(\frac{\partial\alpha}{\partial T}\right)_v \left(\frac{\partial T}{\partial v}\right)_s$$

$$= \left(\frac{\partial\alpha}{\partial v}\right)_T - (\gamma-1)\left(\frac{\partial\alpha}{\partial T}\right)_v \left(\frac{\partial T}{\partial v}\right)_p$$

$$= \left(\frac{\partial\alpha}{\partial v}\right)_T + (\gamma-1)\left(\frac{\partial\alpha}{\partial T}\right)_v \left(\frac{\partial p}{\partial v}\right)_T \Big/ \left(\frac{\partial p}{\partial T}\right)_v, \tag{13}$$

Eq. (12) becomes

$$\left(\frac{\partial\bar{\alpha}}{\partial v}\right)_s^0 = \frac{1}{2}\left(\frac{\partial\alpha}{\partial v}\right)_T^0 + \frac{(\gamma-1)}{2}\left(\frac{\partial p}{\partial v}\right)_T^0 \left(\frac{\partial\alpha}{\partial T}\right)_v^0 \Big/ \left(\frac{\partial p}{\partial T}\right)_v^0. \tag{14}$$

Higher order coefficients may be computed in a similar manner.

The coefficients in Eq. (9) are given in Eqs. (1) and (14) in a form that can be computed from thermodynamic data in which v and T are the independent variables. For data in which p and T are the independent variables, the following equations apply:

$$\alpha = -\frac{v}{p} \Big/ \left(\frac{\partial v}{\partial p}\right)_s, \tag{15}$$

$$\alpha = -\frac{v\gamma}{p} \Big/ \left(\frac{\partial v}{\partial p}\right)_T, \tag{16}$$

$$\bar{\alpha} = \frac{\ln p/p_0}{\int_{p_0}^p \left(\frac{1}{\alpha p}\right) dp}, \tag{17}$$

$$p_0 v_0^{\bar{\alpha}} = p v^{\bar{\alpha}}, \tag{18}$$

$$\bar{\alpha} = \bar{\alpha}^0 + \left(\frac{\partial\bar{\alpha}}{\partial p}\right)_s^0 (p-p_0) + \frac{1}{2!}\left(\frac{\partial^2\bar{\alpha}}{\partial p^2}\right)_s^0 (p-p_0)^2 + \cdots, \tag{19}$$

$$\bar{\alpha}^0 = \alpha^0,$$

$$\left(\frac{\partial\bar{\alpha}}{\partial p}\right)_s^0 = \frac{1}{2}\left(\frac{\partial\alpha}{\partial p}\right)_T^0 - \frac{\gamma-1}{2\gamma}\left(\frac{\partial\alpha}{\partial T}\right)_p^0 \left(\frac{\partial v}{\partial p}\right)_T^0 \Big/ \left(\frac{\partial v}{\partial T}\right)_p^0. \tag{20}$$

Precise evaluation of the difference between $\bar{\alpha}$ and γ requires consistent and accurate equation of state data and specific heat data.

An estimate of the discrepancy can be easily made for the case of an infinitesimal expansion of a gas as may be found in a sound wave.

If the equation of state is taken as

$$pv = RT + f(v, T), \tag{21}$$

the relation between the effective "gamma" and the specific heat ratio is given by

$$\bar{\alpha}^0=\alpha^0=\gamma\left[1-\left(\frac{\partial f}{\partial v}\right)_T\Big/p\right]. \qquad (22)$$

At low pressures, the fractional difference between γ and α^0 is equal to the magnitude of the departure from the ideal gas law, $pv/RT-1$ (for a gas like nitrogen this amounts to a few tenths of one percent), while at high pressures it may be considerably more.

For a finite expansion at moderate initial pressures the fractional difference between γ and $\bar{\alpha}$ has been estimated to be also of the same order of magnitude.

The quantity α^0 is incidentally useful in computing the velocity of sound. From the definition of the velocity of sound it can be shown that

$$C^2=-v^2\left(\frac{\partial p}{\partial v}\right)_s=\alpha^0 pv. \qquad (23)$$

On the Mixing Properties of Non-Linear Condensers

A. VAN DER ZIEL*
N. V. Philips Gloeilampenfabrieken, Eindhoven, Netherlands
(Received March 26, 1948)

The theory of a mixer circuit, containing a non-linear condenser as a mixing element, is developed. It is shown that such a condenser has an imaginary conversion transconductance and a capacitive input and output impedance. It is found that the circuit has widely different properties, depending upon the choice of the frequency, ω_m, of the local oscillator, the i-f frequency, ω_0, and the input frequency, ω_i. The following three cases are considered: (a) $\omega_m=\omega_i+\omega_0$, (b) $\omega_m=\omega_i-\omega_0$, and (c) $\omega_m=\omega_0-\omega_i$. A non-linear condenser in a mixing circuit acts as a transformer; in the cases (b) and (c) the mixing condenser transforms an i-f impedance into a positive input load; in case (a) an i-f impedance is transformed into a negative input load, so that oscillations may even occur. In case (c) the power gain is larger than unity, the power delivered to the circuit by the antenna is smaller than the power dissipated by the output load, and the difference is due to the power delivered by the local oscillator. In case (b) the power gain is smaller than unity, because in this case power is dissipated by the local oscillator. In case (c) instability may occur; the circuit is then capable of splitting the local-oscillator signal into two signals of frequencies, ω_i and ω_0, respectively. The band width and the noise factor of the circuit are also discussed; it is shown that the circuit might have a very low noise factor. A few experiments are given which show qualitative agreement with theory.

I. INTRODUCTION

IN normal mixing circuits non-linear elements which are either formed by a non-linear transductance or by a non-linear resistance (diode, hexode, crystal diode) are used. In such a circuit a small h-f voltage, $V_i=V_i^0\cos(\omega_i t+\varphi_i)$, and a large local-oscillator voltage, $V_m=V_m^0\times\cos\omega_m t$, are applied to the non-linear device and, as a result of the non-linearity, currents of frequency ω_0 will arise, such that:

$$\omega_0=\pm\omega_i\pm n\omega_m. \qquad (1)$$

If this current passes through a circuit tuned at the i-f frequency ω_0, it gives rise to a voltage $V_0=V_0^0\cos(\omega_0 t+\varphi_0)$ across this circuit.** In most cases $n=1$; after (1) there are three possibilities:

(a) $\omega_m=\omega_i+\omega_0$ or $\omega_i=\omega_m-\omega_0$ and $\omega_0=\omega_m-\omega_i$;
(b) $\omega_m=\omega_i-\omega_0$ or $\omega_i=\omega_m+\omega_0$ and $\omega_0=\omega_i-\omega_m$;
(c) $\omega_m=\omega_0-\omega_i$ or $\omega_i=\omega_0-\omega_m$ and $\omega_0=\omega_i+\omega_m$.

The difference between cases (b) and (c) is that the frequencies ω_i and ω_0 are interchanged.

In the case of a diode or a crystal diode the input voltage V_i, the local-oscillator voltage V_m, and the output voltage V_0 are applied to the non-linear element in series, so that the total voltage across this element becomes (compare Fig. 1):

$$V=V_m^0\cos\omega_m t+V_i^0\cos(\omega_i t+\varphi_i)+V_0^0\cos(\omega_0 t+\varphi_0). \qquad (2)$$

It is sometimes useful to change over to the complex notation, but after Haantjes and Tellegen[1]

* Now at the University of British Columbia, Vancouver, Canada.

** We introduce phase angles φ_i and φ_0 into the h-f and i-f voltages, because these voltages are measured across tuned circuits; the phase angles of these voltages may be arbitrarily changed by detuning these circuits.

[1] J. Haantjes and B. D. H. Tellegen, Tijdschrift Ned. Radio Gen. **10**, 237 (1943); Philips Research Reports **2**, 401 (1947).

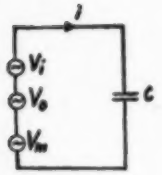


FIG. 1. Input e.m.f. V_i, output e.m.f. V_0, and local oscillator e.m.f. V_m are connected to a non-linear condenser C.

this has to be carried out with some caution—otherwise large errors may result.

It is well known, that any type of non-linear element might be used for mixing. As non-linear dielectrics are now available (e.g., barium titanates), non-linear condensers might be used for this purpose. The theory of mixing circuits containing non-linear condensers has never been worked out in great detail, however. An analysis shows that such circuits have very interesting properties; the results obtained are dealt with in this paper. The theory given here is completely analogous to Haantjes' and Tellegen's analysis of diode mixer circuits. The theory was partly verified by experiments on barium-titanate condensers; experiments are now in progress to verify the remaining part of the theory.

In our analysis we apply the total voltage V of Eq. (2) to the non-linear condenser. The charge Q of the condenser is a function of the applied voltage V and may be developed into a Taylor series after V:

$$Q=\alpha_0 V+\alpha_1 V^2+\alpha_2 V^3+\cdots. \tag{3}$$

For normal condensers $\alpha_1=\alpha_2=\cdots=0$; for non-linear condensers, however, α_1 and α_2 will generally differ from zero. If the voltage V is not too large, it is sufficient to take into account the quadratic term in (3) only; for very high voltages, however, higher order terms in (3) have to be taken into account too.

II. CONVERSION TRANSCONDUCTANCE

It will be shown in this section that a non-linear condenser in a mixing circuit has an imaginary conversion transconductance and a capacitive input and output impedance.

The current i flowing to the condenser is given by

$$i=dQ/dt,$$

which may be written as:***†

$$i=\alpha_0(dV/dt)+2\alpha_1 V(dV/dt)+3\alpha_2 V^2(dV/dt)+\cdots. \tag{5}$$

We now have to introduce (2) into (5). The first term in (5) leads to a capacitive current. The higher order terms lead to mixing; they give rise to frequencies $(\pm n_1\omega_1\pm n_2\omega_0\pm n_3\omega_m)$. Some simplification may be obtained by taking into account that V_i^0 and V_m^0 are very small. For this reason only those frequencies having $n_1=0$, $n_2=1$; $n_1=1$, $n_2=0$; $n_1=n_2=1$ are of interest. The third and higher terms in (5) give rise to mixing on higher harmonics. The higher order terms in (5) only give rise to a change in capacity or in conversion transconductance, but they do not qualitatively change our conclusions, so that it is sufficient to take into account the first two terms of i only. The first term in (5) yields

$$\omega_i\alpha_0 V_i^0\cos(\omega_i t+\varphi_i+\pi/2)+\omega_0\alpha_0 V_0^0\cos(\omega_0 t+\varphi_0+\pi/2)+\omega_m V_m^0\cos(\omega_m t+\pi/2), \tag{6}$$

which shows that for the three signals the condenser has a capacity

$$C'=\alpha_0. \tag{6a}$$

The second term in (5) is more complicated:

$$\begin{aligned}2\alpha_1 V(dV/dt)&=-2\alpha_1[V_m^0\cos\omega_m t+V_i^0\cos(\omega_i t+\varphi_i)\\&\quad+V_0^0\cos(\omega_0 t+\varphi_0)]\\&\quad\cdot[\omega_m V_m^0\sin\omega_m t+\omega_i V_i^0\sin(\omega_i t+\varphi_i)\\&\qquad+\omega_0 V_0^0\sin(\omega_0 t+\varphi_0)].\end{aligned} \tag{7}$$

As we are only interested in mixing and not in frequency multiplication, we have only taken into account the terms having frequencies $(\omega_i+\omega_m)$; $(\omega_i-\omega_m)$; $(\omega_0+\omega_m)$; $(\omega_0-\omega_m)$; $(\omega_i+\omega_0)$;

*** I am indebted to Professor Tellegen for pointing out to me that the theory might be put in the general form in which it is presented here.

† Our theory is analogous to Haantjes' and Tellegen's theory (see reference 1) of the diode; Professor Tellegen remarked that the analogy would be still closer if we put:

$$Q=A_0+A_1\cos\omega_m t+\cdots+(\beta_0+\beta_1\cos\omega_m t+\cdots)v+(\gamma_0+\gamma_1\cos\omega_m t+\cdots)v^2+\cdots,$$

in which:

$$v=V_i^0\cos(\omega_i t+\varphi_i)+V_0^0\cos(\omega_0 t+\varphi_0).$$

This would lead to the following equation:

$$i=\beta_0(dv/dt)+\beta_1(d/dt)(\cos\omega_m t\cdot v).$$

The final results are, of course, the same as in the following method.

$(\omega_i-\omega_0)$. All the other terms in (7) do not take part in the mixing process and will be omitted for the sake of simplicity. Hence

$$\begin{aligned}2\alpha_1 V(dV/dt)&=-\alpha_1 V_i^0 V_m^0[(\omega_i+\omega_m)\sin\{(\omega_i+\omega_m)t+\varphi_i\}\\&\quad+(\omega_i-\omega_m)\sin\{(\omega_i-\omega_m)t+\varphi_i\}]+\\&\quad-\alpha_1 V_0^0 V_m^0[(\omega_0+\omega_m)\sin\{(\omega_0+\omega_m)t+\varphi_0\}\\&\quad+(\omega_0-\omega_m)\sin\{(\omega_0-\omega_m)t+\varphi_0\}]+\\&\quad-\alpha_1 V_i^0 V_0^0[(\omega_0+\omega_i)\sin\{(\omega_0+\omega_i)t+\varphi_0+\varphi_i\}\\&\quad+(\omega_0-\omega_i)\sin\{(\omega_0-\omega_i)t+\varphi_0-\varphi_i\}].\end{aligned}$$

In cases (a), (b), and (c) only those terms have to be taken into account which give rise to frequencies ω_m, ω_i, and ω_0; the other terms do not take any part in the mixing process and are therefore omitted. In case (a) $(\omega_i+\omega_0)=\omega_m$; $(\omega_m-\omega_0)=\omega_i$ and $(\omega_m-\omega_i)=\omega_0$, so that

$$\begin{aligned}2\alpha_1 V(dV/dt)&=\alpha_1[\omega_0 V_i^0 V_m^0\cos(\omega_0 t+\pi/2-\varphi_i)\\&\quad+\omega_i V_0^0 V_m^0\cos(\omega_i t+\pi/2-\varphi_0)\\&\quad+\omega_m V_i^0 V_0^0\cos(\omega_m t+\pi/2+\varphi_i+\varphi_0)].\end{aligned}\tag{7a}$$

In case (b) $(\omega_i-\omega_0)=\omega_m$; $(\omega_m+\omega_0)=\omega_i$, and $(\omega_i-\omega_m)=\omega_0$ so that

$$\begin{aligned}2\alpha_1 V(dV/dt)&=\alpha_1[\omega_0 V_i^0 V_m^0\cos(\omega_0 t+\pi/2+\varphi_i)\\&\quad+\omega_i V_0^0 V_m^0\cos(\omega_i t+\pi/2+\varphi_0)\\&\quad+\omega_m V_i^0 V_0^0\cos(\omega_m t+\pi/2+\varphi_i-\varphi_0)].\end{aligned}\tag{7b}$$

In case (c) $(\omega_0-\omega_i)=\omega_m$; $(\omega_0-\omega_m)=\omega_i$, and $(\omega_i+\omega_m)=\omega_0$, so that

$$\begin{aligned}2\alpha_1 V(dV/dt)&=\alpha_1[\omega_0 V_i^0 V_m^0\cos(\omega_0 t+\pi/2+\varphi_i)\\&\quad+\omega_i V_0^0 V_m^0\cos(\omega_i t+\pi/2+\varphi_0)\\&\quad+\omega_m V_i^0 V_0^0\cos(\omega_m t+\pi/2+\varphi_0-\varphi_i)].\end{aligned}\tag{7c}$$

The first term in (7a), (7b), and (7c) gives the current of frequency ω_0 resulting from the mixing of the signals V_i and V_m. The second term gives the current of frequency ω_i caused by the mixing of the signals V_0 and V_m. These two terms have to be used in order to calculate the conversion transconductances. The third term in (7a), (7b), and (7c) has to be taken into account if the power delivered by the local oscillator is calculated (Section V).

We now change over to the complex notation and try to find the four-pole equations of this mixer circuit. In (7a), (7b), and (7c) we denote the current of frequency ω_i by i_i' and the current of frequency ω_0 by i_0', and we introduce the conversion transconductances, g_{i0}, from input to output and g_{0i} from output to input. In the cases (b) and (c) this can be done in the usual way:

$$i_0'=g_{i0}V_i;\quad i_i'=g_{0i}V_0;\tag{8}$$

but in case (a) due account has to be taken of the fact that in V_i and V_0 the phase angles φ_i and φ_0 occur with a positive sign, whereas (according to (7a)) in i_0' and i_i' they occur with a negative sign. Therefore we put

$$i_0'=g_{i0}V_i^*;\quad i_i'=g_{0i}V_0^*,\tag{8a}$$

where the asterisk denotes the complex conjugate (the signals V_i and V_0 have to be used with phase angles $-\varphi_i$ and $-\varphi_0$). After these definitions and (7a), (7b), and (7c):

$$g_{i0}=j\omega_0\alpha_1 V_m^0;\quad g_{0i}=j\omega_i\alpha_1 V_m^0,\tag{8b}$$

for all three cases. Denoting the total currents of frequencies ω_i and ω_0 by i_i and i_0, respectively, the four-pole equations of the circuit are:

$$\begin{aligned}i_i&=j\omega_i C' V_i+g_{0i}V_0^*,\\i_0&=g_{i0}V_i^*+j\omega_0 C' V_0,\end{aligned}\tag{9a}$$

$$\begin{aligned}i_i&=j\omega_i C' V_i+g_{0i}V_0,\\i_0&=g_{i0}V_i+j\omega_0 C' V_0,\end{aligned}\tag{9b, 9c}$$

which are completely similar to the equations for diode mixers as given by Haantjes and Tellegen (compare also Torrey and Whitmer[2]).

The above theory shows that the conversion transconductance is purely imaginary and that it is proportional to the i-f frequency. (In case of large dielectric or hysteresis losses this does not hold.) Hence

$$g_{i0}=j|g_{i0}|;\quad g_{0i}=j|g_{0i}|.\tag{10}$$

Measurements were carried out on a barium titanate condenser, having a capacity of about 30 $\mu\mu$f. We found for mixing from 8.5 to 20 m a conversion transconductance of 65 μa/v, and for mixing from 20 to 6.7 m a conversion transconductance of 160 μa/v (V_m^0 was about 10v). The ratio of the transconductances is about 2.5 and the ratio of the i-f frequencies is 3.0, so that (8b) is approximately verified. We did not succeed in measuring the conversion transconductance of Rochelle salt condensers.

[2] H. C. Torrey and Ch. A. Whitmer, *Crystal Rectifiers* (M.I.T. Rad. Lab. Series) (McGraw-Hill Book Company, Inc., New York, 1948), Vol. 15.

III. TRANSFORMER PROPERTIES OF NON-LINEAR MIXING CONDENSERS

It will be shown in this section that a non-linear condenser in a mixing circuit acts as a transformer. It will be shown that in the cases (b) and (c) the mixing condenser transforms an i-f impedance into a positive input load; in case (a) an i-f impedance is transformed into a negative input load, so that oscillations may even occur.

We assume that an antenna is coupled to the h-f input circuit and that a resistance R_0 is connected to the i-f output circuit of the mixer (Fig. 2). Let the antenna output be represented by an e.m.f. V_a in series with a resistance R_a, such that the available antenna power, $P_a = \frac{1}{4}V_a^2/R_a$, has a fixed value. R_a and R_0 are assumed to be variable; their values have to be chosen such that a maximum amount of power is dissipated by R_0. We assume also that the input circuit is tuned at the frequency ω_i and the output circuit at the frequency ω_0, that the resonance impedance of the circuits is very high in comparison to R_a and R_0, respectively, and that the output circuit has zero impedance for the frequencies ω_i and ω_m and the input circuit zero impedance for the frequencies ω_0 and ω_m. In order to represent the tuning of the input and output circuits a (negative) capacity C_i is connected in parallel to the input circuit, and a (negative) capacity C_0 in parallel to the output circuit. Tuning means that

$$C_i + C' = 0; \quad C_0 + C' = 0. \tag{11}$$

We therefore have to introduce into (9a), (9b), and (9c):

$$\begin{aligned} i_0 &= -V_0(1/R_0 + j\omega_0 C_0), \\ i_i &= (V_a - V_i)/R_a - j\omega_i C_i V_i \\ &= V_a/R_a - (1/R_a + j\omega_i C_i)V_i \end{aligned} \tag{12}$$

(compare Fig. 2).

We first discuss the cases (b) and (c). Introducing (12) into (9b) and (9c) and taking into account that the input and output circuit are tuned, we obtain

$$\begin{aligned} V_a/R_a &= 1/R_a V_i + j|g_{i0}|V_0, \\ 0 &= j|g_{i0}|V_i + 1/R_0 V_0. \end{aligned} \tag{13}$$

Hence

$$V_0 = -j|g_{i0}|R_0 V_i, \tag{14}$$

$$V_i = V_a/R_a \cdot [1/R_a + |g_{i0}g_{0i}|R_0]^{-1}. \tag{15}$$

Introducing this into (12) we obtain

$$\begin{aligned} i_i &= V_i[j\omega_i C' + |g_{i0}g_{0i}|R_0] \\ &= V_i[1/R_i' + j\omega_i C'], \end{aligned} \tag{16}$$

which shows that the input impedance consists of the capacity C', as should be expected, connected in parallel to a resistance R_i':

$$1/R_i' = |g_{i0}g_{0i}|R_0. \tag{16a}$$

In the same way it may be shown that the antenna resistance R_a is transformed into an output load resistance R_0':

$$1/R_0' = |g_{i0}g_{0i}|R_a. \tag{16b}$$

Hence the mixer circuit acts as a transformer, transforming an input load into an output load an output load into an input load. $R_i' = \infty$ if $R_0 = 0$ and $R_i' = 0$ if $R_0 = \infty$, whereas $R_0' = \infty$ if $R_a = 0$ and $R_0' = 0$ if $R_a = \infty$. The maximum amount of energy is transferred if $R_i' = R_a$ and $R_0' = R_0$; in this case the antenna is matched to the input and the output load R_0 is matched to the output of the circuit.

We now turn to case (a). Introducing (12) into (9a) and bearing in mind that the input and output circuit are tuned, we obtain

$$\begin{aligned} V_a/R_a &= 1/R_a V_i + j|g_{i0}|V_0^*, \\ 0 &= j|g_{i0}|V_i^* + 1/R_0 V_0; \end{aligned} \tag{17}$$

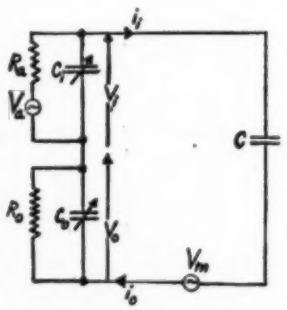


FIG. 2. Mixing circuit containing a non-linear condenser C. An antenna represented by an e.m.f. V_a and an internal resistance R_a is connected to the input circuit; a load R_0 is connected to the output circuit. A local-oscillator voltage is also applied. The input and output circuits are tuned; this is represented by the capacities C_i and C_0; their values have to be chosen such that the input is tuned to the frequency ω_i and the output to the frequency ω_0. The input voltage is denoted by V_i, the input current by i_i, the output voltage by V_0, the output current by i_0.

or

$$0=-j|g_{i0}|\,V_i+1/R_0V_0^*,$$

which is obtained by substituting j by $-j$ in the second equation. Hence

$$V_0^*=j|g_{i0}|\,R_0V_i, \tag{18}$$

$$V_i=V_a/R_a[1/R_a-|g_{0i}g_{i0}|\,R_0]^{-1}. \tag{19}$$

Introducing this into (12) we obtain

$$i_i=V_i[j\omega_iC'-|g_{0i}g_{i0}|\,R_0] \\ =V_i[1/R_i''+j\omega_iC'], \tag{20}$$

which shows that the input impedance consists of the capacity C', as should be expected, in parallel to a negative resistance R_i'', such that

$$1/R_i''=-|g_{0i}g_{i0}|\,R_0. \tag{20a}$$

In the same way it may be shown that the antenna resistance R_a is transformed into an output load resistance R_0'':

$$1/R_0''=-|g_{0i}g_{i0}|\,R_a. \tag{20b}$$

$R_i''=\infty$ if $R_0=0$, $R_i''=0$ if $R_0=\infty$, $R_0''=\infty$ if $R_a=0$ and $R_0''=0$ if $R_a=\infty$. In this case it is impossible to match the antenna to the input or the output load, R_0, to the output of the circuit. Under favorable conditions oscillations may occur. This is the case if

$$1/R_a+1/R_i''\leqq 0 \quad \text{or} \quad 1/R_0+1/R_0''\leqq 0. \tag{20c}$$

After (20a) and (20b) both conditions are satisfied simultaneously, which is not surprising, for if the circuit generates one of the frequencies ω_0 and ω_i, it will automatically generate the other one in the mixing process.

IV. POWER GAIN

It will be shown in this section that under certain conditions the available output power of the mixer may be larger than the available antenna power. Moreover, the band width of the mixer circuit will be discussed.

We first discuss the power gain in the cases (b) and (c). After (14) and (15) the power delivered to the resistance R_0 is

$$|V_0|^2/R_0=|g_{i0}|^2R_0V_i^2 \\ =|g_{i0}|^2R_0V_a^2/[1+|g_{0i}g_{i0}|\,R_0R_a]^2. \tag{21}$$

Dividing this by the available antenna power, $P_a=\frac{1}{4}V_a^2/R_a$, we obtain for the power gain g:

$$g=4R_0R_a|g_{0i}g_{i0}|/[1+|g_{0i}g_{i0}|\,R_0R_a]^2 \\ \cdot|g_{i0}|/|g_{0i}|, \tag{22}$$

which attains a maximum value

$$g_{\max}=|g_{i0}|/|g_{0i}| \quad \text{for} \quad R_0R_a|g_{0i}g_{i0}|=1. \tag{22a}$$

This shows that a gain >1 is possible, if $|g_{i0}|>|g_{0i}|$. This is the case if $\omega_0>\omega_i$, because after (8b):

$$g_{\max}=\omega_0/\omega_i. \tag{22b}$$

We now turn to case (a). After (18) and (19) the power delivered to the resistance R_0 is

$$|V_0|^2/R_0=|V_0^*|^2/R_0=|g_{i0}|^2R_0V_i^2 \\ =|g_{i0}|^2R_0V_a^2/[1-|g_{0i}g_{i0}|\,R_0R_a]^2. \tag{23}$$

The power gain is obtained by dividing this expression by the available antenna power $P_a=\frac{1}{4}V_a^2/R_a$:

$$g=4R_0R_a|g_{0i}g_{i0}|/[1-|g_{0i}g_{i0}|\,R_0R_a]^2 \\ \cdot|g_{i0}|/|g_{0i}|. \tag{24}$$

The power gain becomes infinitely large if R_0 and R_a are properly chosen; this means that oscillations may occur. The power gain is largest if $|g_{i0}|>|g_{0i}|$. This is the case if $\omega_0>\omega_i$. The large values of the power gain are promising, but the band width of the circuit has to be considered too, and the circuit will be useless if the band width is too small.

For a fair comparison between the cases (b) and (c) on the one hand, and case (a) on the other hand, the band width has to be the same in both cases. The input and output circuit capacity will generally be formed by the capacity C' of the condenser itself. As the band width is determined by the total impedance of the tuned circuit and by the circuit capacity, and the band width of the input and the output circuit should be equal, R_0 should be chosen equal to R_a in both cases. Putting by definition:

$$|g_{0i}g_{i0}|=R^{-2}, \tag{25}$$

$$R_0=R_a=R'' \text{ in case (a), and} \tag{25a}$$

$$R_0=R_a=R' \text{ in cases (b) and (c),} \tag{25b}$$

we have for the power gain g in relative units:

$$g=4(R''/R)^2/[1-(R''/R)^2]^2 \text{ in case (a),} \tag{26a}$$

$$g=4(R'/R)^2/[1+(R'/R)^2]^2 \\ \text{in cases (b) and (c).} \tag{26b}$$

If the total impedance of the tuned input circuit (being equal to the impedance of the tuned output circuit) is denoted by R_1 for cases (b) and (c) and by R_2 for case (a), we have

$$1/R_2=1/R''-R''/R^2 \text{ in case (a)}, \quad (27a)$$

$$1/R_1=1/R'+R'/R^2 \text{ in cases (b) and (c)} \quad (27b)$$

after (16a) and (20a). For equal band width R_1 and R_2 should be equal. $R_1=R_2$ means

$$1/R''-R''/R^2=1/R'+R'/R^2,$$

or, by multiplying both sides of this equation by R,

$$R/R''[1-(R''/R)^2]=R/R'[1+(R'/R)^2],$$

or

$$[1-(R''/R)^2]^2/(R''/R)^2=[1+(R'/R)^2]^2/(R'/R)^2$$

so that after (24a) and (24b) the power gain is the same in both cases as long as equal band widths are considered.

There is, however, some difference between both cases. After (27b) the total impedance of the tuned circuits is a maximum if $R'=R$, and hence the band width then is a minimum. Much smaller band widths and correspondingly larger power gains can be obtained in case (a).

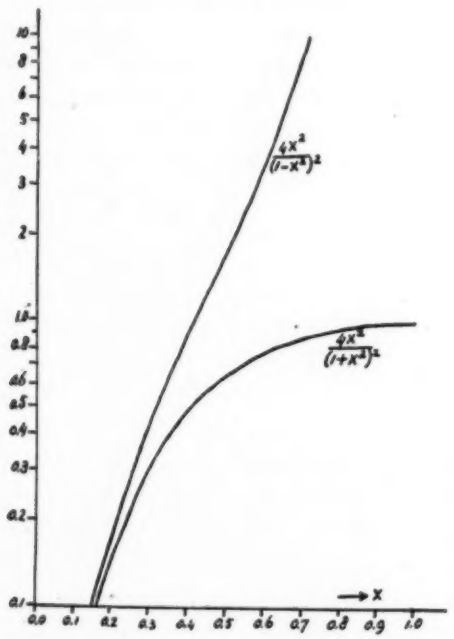


FIG. 3. Power gain g as a function of R_a and R_0. In case (a) $g=4x^2(1-x^2)^{-2}$, where $x=R''/R$; in cases (b) and (c) $g=4x^2(1+x^2)^{-2}$, where $x=R'/R$. $R=|g_{i0}g_{0i}|^{-\frac{1}{2}}$.

This would be especially useful if the value of R were sufficiently small, because in that case a reasonable band width and a large power gain might be obtained simultaneously.

Figures 3 and 4 show the power gain and the band width as a function of R'/R and R''/R; the power gain in case (a) increases much faster with R''/R than the band width decreases. This means that even for $\omega_0 \ll \omega_i$ a reasonable power gain might be obtained by chosing the value of R''/R near to unity.

V. BALANCE OF POWER

A power gain larger than unity means that the available power at the mixer output is larger than the available antenna power. It will be shown that this power is delivered by the local oscillator.

The power P_i delivered to the input generally differs from the power P_0 dissipated by R_0. In case (c) $P_0>P_i$, the difference (P_0-P_i) is delivered by the local oscillator. In case (b) $P_0<P_i$, the difference (P_i-P_0) is dissipated by the local oscillator. Case (a) is still more interesting; a signal is applied by the antenna, but no power is delivered to the circuit; on the contrary, the local oscillator delivers power to the antenna and to the resistance R_0. If the circuit is such that oscillations occur, we then have a device which splits a signal of angular frequency, ω_m, into two signals of frequencies, ω_i and ω_0, respectively, such that $\omega_m=(\omega_i+\omega_0)$.

We first investigate case (c). After (7c) the current i_m of frequency ω_m becomes

$$i_m=\omega_m\alpha_1 V_i^0 V_0^0 \cos(\omega_m t+\pi/2+\varphi_0-\varphi_i).$$

The power delivered by the local oscillator is the average value of $i_m V_m$:

$$P_m=\langle i_m V_m\rangle_{Av}=-\tfrac{1}{2}\omega_m\alpha_1 V_m^0 V_i^0 V_0^0 \sin(\varphi_0-\varphi_i).$$

After (14) and (15) $\varphi_i=0$ and $\varphi_0=-\pi/2$ for tuned circuits; hence

$$P_m=\tfrac{1}{2}\omega_m\alpha_1 V_m^0 V_i^0 V_0^0=\tfrac{1}{2}\omega_0\omega_m\alpha_1^2 V_m^{02} V_i^{02} R_0 \quad (28)$$

after (14). After (21) the power delivered to the output circuit is

$$P_0=\tfrac{1}{2}V_0^{02}/R_0=\tfrac{1}{2}\omega_0^2\alpha_1^2 V_m^{02} V_i^{02} R_0, \quad (28a)$$

and the power delivered by the antenna to the

input circuit is

$$P_i=\tfrac{1}{2}V_i^{02}/R_i''=\tfrac{1}{2}\omega_i\omega_0\alpha_1^2V_m^{02}V_i^{02}R_0 \quad (28b)$$

after (16a) and (8b). Because $\omega_0=(\omega_i+\omega_m)$, we have in this case

$$P_0=P_i+P_m, \quad (28c)$$

so that the power dissipated by R_0 is partly delivered by the antenna and partly by the local oscillator.

In case (b) P_0 and P_i remain the same, but P_m changes sign, because after (7b) the current i_m is in this case:

$$i_m=\omega_m\alpha_1V_i^0V_0^0\cos(\omega_m t+\pi/2-\varphi_0+\varphi_i).$$

After (14) and (15) $\varphi_i=0$ and $\varphi_0=-\pi/2$. Calculating P_m in the same way as in case (c), we have

$$P_m=-\tfrac{1}{2}\omega_0\omega_m\alpha_1^2V_m^{02}V_i^{02}R_0. \quad (29)$$

As $(\omega_0+\omega_m)=\omega_i$, we obtain

$$P_m=-(P_i-P_0), \quad (29a)$$

which shows that the difference between input power and power dissipated by R_0 is dissipated by the local oscillator.

We finally turn to case (a). After (7a),

$$i_m=\omega_m\alpha_1V_i^0V_0^0\cos(\omega_m t+\pi/2+\varphi_0+\varphi_i),$$

so that

$$P_m=\langle i_mV_m\rangle_{Av}=-\tfrac{1}{2}\omega_m\alpha_1V_m^0V_i^0V_0^0\sin(\varphi_0+\varphi_i).$$

After (19) $\varphi_i=0$, whereas after (18) V_0^* has a phase angle of $+\pi/2$ and hence V_0 a phase angle of $-\pi/2$, so that after (18)

$$P_m=\tfrac{1}{2}\omega_0\omega_m\alpha_1^2V_m^{02}V_i^{02}R_0, \quad (30)$$

as in case (c).

The power dissipated by R_0 is again given by (28a), but the input power is now

$$P_i=\tfrac{1}{2}V_i^{02}/R_i''=-\tfrac{1}{2}\omega_i\omega_0\alpha_1^2V_m^{02}V_i^{02}R_0 \quad (30a)$$

according to (20a); this means that power is delivered to the antenna. Hence we have

$$P_m=P_0+(-P_i), \quad (30b)$$

which shows that the power delivered by the local oscillator is partly dissipated by the output, whereas the amount of power $(-P_i)$ is dissipated by the antenna. After (30a) and (28a),

$$P_0:(-P_i)=\omega_0:\omega_i, \quad (30c)$$

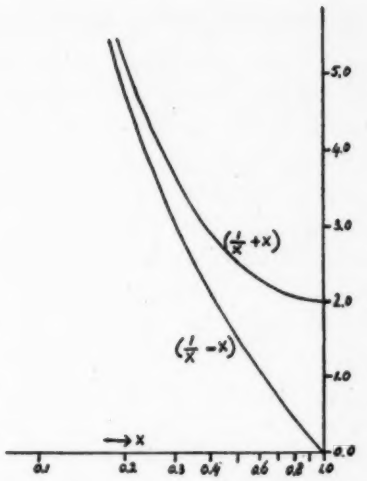


FIG. 4. Total impedance of the tuned input and output circuits as a function of R_a and R_0. In case (a) we have: $R/R_2=(1/x-x)$, where $x=R''/R$, R_2 is the impedance of the tuned circuits, and $R''=R_a=R_0$. In cases (b) and (c) we have $R/R_1=(1/x+x)$, where $x=R'/R$, R_1 is the impedance of the tuned circuits, and $R'=R_a=R_0$. $R=|g_{0i}g_{i0}|^{-\frac{1}{2}}$.

so that the power is delivered in the ratio of the frequencies ω_0 and ω_i. If the circuit is stable, the total amount of power delivered by the local oscillator depends upon V_i^0 and hence upon V_a^0; if oscillations occur no antenna signal is needed, and our assumption that V_0 and V_i are small quantities is no longer valid.

It was assumed that the antenna was matched to the input and the resistance R_0 to the output in the cases (b) and (c), whereas in all three cases the input and output circuits were assumed to be tuned. It may be shown easily, that the relations (28c), (29a), and (30b) also hold if these conditions are not satisfied.

As the measured values of the conversion transconductance had the order of magnitude of a hundred μa/v, R should have the order of magnitude of 10000Ω, so that it should be comparatively easy to obtain oscillations. A similar phenomenon is described by North,[3] who found that with welded contact germanium rectifiers power gain could be obtained and that an i-f circuit could oscillate at the resonant frequency of the circuit, at least up to 34 megacycles. This was ascribed to the fact that the contact capacity varied with bias.

[3] H. Q. North, J. App. Phys. **17**, 912 (1946).

VI. NOISE CONSIDERATIONS

If thermal noise were present only, one would expect the circuit to have a very low noise factor. If dielectric losses (and hysteresis losses) were negligible the circuit would not contribute to the noise of the amplifier at all; its only function would be to amplify the antenna signal and the antenna noise. Hence, if the noise factor of the i-f part of the receiver is denoted by F and the power gain of the mixer circuit by g, we would have

$$F'=1+(F-1)/g, \tag{31}$$

if F' is the noise factor of the complete receiver. As g may be larger than unity, the noise factor of the complete receiver might be much smaller than the noise factor of the i-f part.

The power gain is larger than unity if $\omega_0>\omega_i$. As mixing is generally carried out with the intention to change over to a lower frequency, this does not seem very useful. But the mixing circuit described here has the property to increase the power without introducing extra noise. A crystal diode might then be used to mix from the high first i-f frequency to a much lower second if-frequency. Because of the increase in power caused by the first mixer stage, the noise figure of the receiver might be much better than in the case of direct mixing to the lower i-f frequency.

Other types of noise might also be present, *viz.*, noise caused by spontaneous fluctuations in the polarization of the dielectric and noise caused by the fact that for variable voltages the polarization might show a large number of small abrupt changes, similar to the *Barkhausen* effect in ferromagnetic substances.

The spontaneous fluctuations of polarization would give rise to spontaneous fluctuations in capacity and as a result of the applied voltage this would give rise to extra noise. It is a phenomenon similar to the extra noise in carbon resistors through which a direct current is passing. Nothing is known about the magnitude of this effect. The same holds for the "*Barkhausen* effect" in these dielectrics. These problems are still under investigation and we hope to report on them in the near future.

VII. COMPARISON WITH THE WORK ON GERMANIUM CRYSTAL DIODES

Since this paper was written, a fuller account of the work on welded contact Germanium crystal diodes, first described by North,[3] was published in the Massachusetts Institute of Technology Radiation Laboratory Series.

Pound[4] gives many interesting details, e.g., about the measured power gain and the measured negative i-f conductance. Torrey and Whitmer[2] give a thorough theoretical discussion of the problem in volume **15**, Chapter 5.13 and 13 of that series. Their results are in good agreement with the results presented here. In their case the conversion transconductances are not purely imaginary but contain a real and an imaginary part, the imaginary part being due to the variable capacitance of the barrier layer in the crystal diode. They investigate under which conditions a power gain can be obtained and show how the variable capacitance of the barrier layer can give rise to a negative i-f conductance. They also discuss the effect of the image side band.

In some respects our work is more general, as it gives a unified treatment of the problem and discusses energy gain and input and output impedances in terms of the two conversion transconductances, g_{i0} and g_{0i}. In other respects it is less general, because the influence of the real components of the conversion transconductances and the influence of the image side band are not discussed here. The latter would certainly be interesting because it may give rise to a negative output impedance in cases (b) and (c) too.

ACKNOWLEDGMENT

The author wishes to express his thanks to Dr. G. H. Jonker and Dr. J. H. van Santen of N. V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands, for drawing his attention to this problem. They discovered the non-linear properties of barium-titanates and supplied the samples for the measurements. The author is indebted to Professor B. D. H. Tellegen for important improvements in the manuscript.

[4] R. V. Pound, *Microwave Mixers* (M.I.T. Rad. Lab. Series) (McGraw-Hill Book Company, Inc., New York, 1948), Vol. 16.

An Extension of Lagrange's Equations to Electromagnetic Field Problems*

PRESCOTT D. CROUT
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received March 29, 1948)

This paper contains the basic theory of an extension of Lagrange's equations which renders these equations and the procedures associated with them in mechanics applicable to electromagnetic field problems such as are encountered in ultra high frequency work and other branches of engineering. In certain important cases the Lagrangean procedures so obtained give equivalent networks. Successful applications of these procedures to several practical problems are indicated.

1. INTRODUCTION

THE purpose of this paper is to present the basic theory of an extension of Lagrange's equations which renders these equations and the procedures associated with them in mechanics applicable to electromagnetic field problems, such as the echo box, wave guide structures, the structure of the magnetron, and so on.

Assuming that in a given problem the electric field can at each instant be sufficiently approximated by a linear combination of chosen "current modes," the coefficients in which combination are taken as functions of time and of generalized coordinates; it is proved that Lagrange's equations are satisfied if the roles of the kinetic and potential energies are taken by the magnetic and electric energies, respectively.

Next it is shown that in certain cases the generalized coordinates[1] that are used in connection with the Lagrange equations may be considered as the charges which have flowed around the various meshes of a lumped network whose Kirchoff voltage equations are identical with the Lagrange equations.[2] This lumped network is hence equivalent to the given system to the degree of approximation permitted by the choice and number of generalized coordinates. The Lagrangian procedures thus give equivalent networks.

A duality is shown to exist between electric and magnetic quantities, from which it follows that if the magnetic field is sufficiently approximated by a linear combination of "flux modes" and if the coefficients in this combination are taken as functions of time and of generalized coordinates, then Lagrange's equations are satisfied if the roles of the kinetic and potential energies are taken by the electric and magnetic energies, respectively.

It is finally shown that Lagrange's equations are still valid if the chosen current or flux modes change with time in any specified manner. This fact may be useful in problems involving moving coordinate systems, and in propagation problems.

The theory obtained has been applied successfully to a number of problems which arise in the design of echo boxes.[3] The effect of coupling loops, resistance of the walls, deviation of the cross section of the cylindrical tube from circular, deviation of the planes of the ends of the box from perpendicularity to the axis, the presence of an air gap between the piston and the cylindrical wall, eccentricity of the piston, the presence of a back cavity and of bakelite in this cavity were all considered; experimental data confirming the calculated results were obtained. Lagrangian procedures have also been applied to obtain natural frequencies of a torus, and to calculate the behavior of the resistance type

* This paper is based on work done for the Office of Scientific Research and Development under contract OEMsr-262 with the Massachusetts Institute of Technology.

[1] Although the use of Lagrange's equations implies the use of generalized coordinates, the converse is not true. Generalized coordinates can be used in connection with a method of virtual displacements to solve certain types of electromagnetic field problems. See P. D. Crout, "A method of virtual displacements for electrical systems with applications to pulse transformers," Proc. I.R.E. **35**, 1236 (1947).

[2] It is evident that equivalent networks can be devised which are the duals of those just considered, and in which the q's are node voltages instead of mesh charges.

[3] P. D. Crout and N. H. Painter, "A Treatment of Echo Box Problems by Lagrangian Procedures," Radiation Laboratory Reports 629 and 696; A. Banos, Jr., "Theory of ringing time of tunable echo boxes," Radiation Laboratory Report 630.

wave guide attenuator; furthermore, they are being applied to various magnetron problems.

The idea of extending the equations of mechanics to electrical problems is not new. It is obvious that if the charges which have flowed around the various meshes of a lumped network be taken as generalized coordinates, and if the magnetic energy be treated as kinetic energy, then Lagrange's equations give Kirchoff's voltage equations. Similarly if the node voltages of a lumped network be taken as generalized coordinates, and if the electric energy be treated as kinetic energy, Lagrange's equations give Kirchoff's current equations. Lagrange's equations have also been applied to lumped electromechanical systems.

A "dynamical theory of currents" was devised by Maxwell. The currents considered were linear, two sets of generalized coordinates being used. One set served to locate a number of circuits, the other to fix the charges which have flowed around these circuits. The language is hence that of lumped networks, although Maxwell uses one linear circuit to probe the field due to others and thus obtains results in agreement with Faraday's law.

There are a number of variational principles, which differ among themselves in the integral that is rendered stationary, in the quantities that are varied, and in the constraints which are satisfied during the variation. These, however, do not suffice to give the results of the present paper.[4]

2. VIRTUAL FIELDS, VIRTUAL MODES, AND GENERALIZED COORDINATES

Let us consider that part of an electromagnetic field which lies within a closed surface S. In order that this surface can bound the field physically as well as in our imagination we consider it as carrying surface distributions of electric current and magnetic flux given by

$$J_s = \frac{10}{4\pi} H \times 1_n \text{ amp./cm},$$
$$\dot{B}_s = -10^8 E \times 1_n \text{ maxwells/cm sec.}, \tag{1}$$

respectively, where 1_n is a unit vector normal to the surface outward, and where "practical" units are used for later convenience. The surface distributions (1) do not alter the field within the surface but reduce it to zero outside, and provide return paths for the electric current and magnetic flux which enter the surface from within. (See Appendix.) To provide for the physical existence of these surface distributions we proceed as follows.

Let the electric current distribution lie just inside of the magnetic flux distribution, then the latter, being completely shielded by the former, lies in a region of zero magnetic intensity. Without altering the field situation we may therefore consider the surface to be a perfect magnetic conductor, and hence to provide a path for the surface flux. The electric current distribution is completely unshielded, and will hence be considered as passing through a surface layer of generators which provide an e.m.f. per cm equal to the negative of the tangential component of E. The purpose of these generators is the same as that of the single generator at the end of a transmission line, namely to provide for the current a return path over which the electric field is neutralized (by the e.m.f.). If this vanishes, as is the case where the field is bounded by a perfect electrical conductor, we see from (1) that there is no surface flux distribution and hence that it is not necessary to render the surface magnetically conducting. Similarly, if the tangential component of H vanishes, as is the case where the field is bounded by a perfect magnetic conductor, we see from (1) that there is no surface current distribution.

Let us now suppose that the actual electromagnetic field can at each instant be sufficiently approximated by superimposing the fields due to N suitably chosen electric current modes. Each

[4] H. F. Olson, *Dynamical Analogies* (D. Van Nostrand Company, New York, 1943), p. 37; J. C. Maxwell, *A Treatise on Electricity and Magnetism* (Clarendon Press, Oxford, 1892), Vol. 2, p. 199; J. H. Jeans, *The Mathematical Theory of Electricity and Magnetism* (Cambridge University Press, Teddington, 1925), p. 485; G. H. Livens, "On the Fundamental Formulations of Electrodynamics," Phil. Trans. Roy. Soc. (Series A) **220**, 207 (1920); G. H. Livens, *The Theory of Electricity* (Cambridge University Press, Teddington, 1926), pp. 199, 244; H. Bateman, "Sidelights on Electromagnetic Theory," Phys. Rev. **45**, 721 (1934); M. Pastori, "Principi Variazionali del Campo Elettromagnetico," Atti della Reale Accademia Nazionale dei Lincei, **29**, 48, 145 (1939).

mode comprises a current field[5]

$$J_i = \dot{Q}_i(t) f_i(x, y, z) \text{ amp./cm}^2, \qquad i = 1, 2, \cdots, N, \quad (2)$$
$$J_{si} = \dot{Q}_i(t) f_{si}(x, y, z) \text{ amp./cm},$$

in the volume and on the bounding surface, respectively, where f_i and f_{si} are vector point functions and $Q_i(t)$ is a magnification factor. The currents indicated by (2) flow around in closed paths, f_i and f_{si} being chosen so that

$$\nabla \cdot f_i = 0 \text{ at interior points,}$$
$$\nabla_s \cdot f_{si} = 1_n \cdot f_i \text{ at points of the bounding surface,} \quad (3)$$

∇_s being the surface divergence.

It follows that the current field J and J_s obtained by superimposing the fields of the various modes satisfies (3) and has this same property. At each point the nature of the current (conduction or displacement) and the medium if any that carries the current is specified as part of the choice of the mode.

Each mode also comprises the electric field and the magnetic field due to the currents (1), the former being

$$E_i = \frac{4\pi c^2 10^{-9}}{K(x, y, z)} Q_i(t) f_i(x, y, z) \text{ volts/cm}$$
for displacement currents,
$$E_i = \rho(x, y, z) \dot{Q}_i(t) f_i(x, y, z) \text{ volts/cm} \quad (4)$$
for conduction currents;

and the latter being

$$B_i = \dot{Q}_i(t) g_i(x, y, z) \text{ at interior points,}$$
$$B_{si} = \dot{Q}_i(t) g_{si}(x, y, z) \text{ at points on the bounding surface.} \quad (5)$$

Here $\rho(x, y, z)$, $K(x, y, z)$, and $\mu(x, y, z)$ are the resistivity, dielectric constant, and permeability, respectively. From Ampere's law and Stoke's theorem we have

$$\nabla \times \left(\frac{g_i}{\mu}\right) = \frac{4\pi f_i}{10}. \quad (6)$$

As before, we consider the bounding surface to be a perfect magnetic conductor just inside of which is a surface distribution of generators which for the ith mode provide an e.m.f. per cm equal to the negative of the tangential component of E_i. The surface current distribution J_{si} which flows through the generators is put equal to

$$J_{si} = \frac{10}{4\pi} H_i \times 1_n, \quad (7)$$

similar to (1), which satisfies continuity conditions of the form (3), and renders the magnetic intensity zero outside of the current sheet—in particular in the magnetic conducting sheet. (See Appendix.) From (7), (2), and (5) it follows that

$$f_{si} = \frac{10}{4\pi} \left(\frac{g_i}{\mu}\right) \times 1_n. \quad (8)$$

The surface flux B_{si} in the magnetic conducting sheet consists of two parts. One part comprises those tubes which at some point leave the surface, and which form return paths for the flux brought to the surface by the normal component of H_i. The other part comprises those tubes which lie entirely in the surface, and which prevent the electric field E_i from penetrating this surface. An explicit expression for B_{si} is not required in the work that follows. The sum of the B_{si}'s, however, gives a surface flux density B_s which satisfies (1).

As before, we see that where the surface is a perfect electrical conductor B_{si} vanishes, and it need not be a magnetic conductor; also that where the surface is a perfect magnetic conductor J_{si} vanishes and the surface distribution of generators may be omitted. Since modes of the type just described exist alone only in one's imagination and need not be capable of existing alone physically, they will be called "virtual modes."

As an example the field in a cylindrical cavity with a coupling loop can be approximated by a number of current modes whose f_i's and f_{si}'s are given by the current fields of corresponding na-

[5] Surface and curvilinear current distributions may be included by extending the meaning of (2) so that J_i is the volume current density in amp./cm^3, surface current density in amps./cm, or curvilinear current in amp. at point (x, y, z) in a volume, on a surface, or on a curve, respectively. This could have been but was not done in the above derivation of Lagrange's equations, because since surface and curvilinear current distributions may be considered as limiting cases of volume current distributions, it was felt that nothing essential would be added by thus complicating the derivation, in which only the currents in the bounding surface are included explicitly. Later in applying Lagrange's equations the above extension of (2) is used wherever convenient.

tural modes at any instant, together with a mode comprising only surface conduction currents whose corresponding f_{si} is the surface current distribution which is caused by one ampere in the coupling loop and which prevents the magnetic field from penetrating any metallic surface. That part of the boundary surface which cuts the coaxial line from the coupling loop is disk shaped. In accordance with (8) the inner surface of this disk carries a distribution of generators whose e.m.f.'s are radial (away from the central conductor) and add up to the line voltage, and the outer surface carries a surface distribution of magnetic flux which lies entirely in the disk and circulates around the central conductor.

We shall finally suppose that the Q_i's are functions of n generalized coordinates $q_1, q_2, \cdots, q_n$, and time, thus

$$Q_i(t) = Q_i(q_1, q_2, \cdots, q_n, t), \quad i = 1, 2, \cdots, N, \quad (9)$$

where the q's are unknown functions of time. Our problem then becomes that of determining the q's as functions of time. The fact that the q's instead of the Q's are taken as generalized coordinates adds a considerable amount of generality and flexibility to the results that will be obtained. Certain applications of this flexibility are given in Section 9. The idea of having the q's distinct from the Q's was originally suggested by analogy with a mechanics problem in which the Q's determined displacements directly, and in which it was noted that the q's could be but were not necessarily of the nature of displacements.

3. VIRTUAL WORK

Let us consider the given electromagnetic system to be in operation, the q's being suitable but unknown functions of time; and at time t let the system be given a virtual displacement δq_k. By this we mean that at time t time is stopped in the sense that all e.m.f.'s and electric field intensities which existed then are considered as persisting unchanged; and that q_k is increased by an infinitessimal amount δq_k, the t in (9) being held fixed. The variation δq_k causes a virtual field of flow of charge, the word "charge" being used in the sense of time integral of current, which field may be considered as made up of an infinite number of closed, elementary, infinitely thin tubes of flow, the charge flowing in each being proportional to δq_k. On bounding surfaces these tubes become strips. Considering the positive direction to be the direction of flow for δq_k positive, we have for any tube

e.m.f. due to voltage sources (generators)

$$\text{in tube} = \text{Impedance drop} \int_s E \cdot ds$$

$$+ \text{Induced back voltage } 10^{-8} \int_a \dot{B} \cdot da, \quad (10)$$

where ds is a vector element of tube length and da is the vector element of an area which spans the tube. Multiplying (10) by the virtual charge in that tube, adding the corresponding expressions for all the tubes, and dividing by δq_k, we have

$$\frac{\delta W_s}{\delta q_k} = \frac{\delta W_a}{\delta q_k} + \frac{\delta W_b}{\delta q_k}, \quad (11)$$

where δW_s is the total virtual work *done by* the voltage sources, and δW_a and δW_b are the virtual works *done on* the impedance drops and the induced back voltages, respectively. The word "virtual" is used to emphasize the fact that the process is entirely artificial, and can exist only in one's imagination.[6]

[6] Regardless of the number of modes used the nature of the approximate result that will be obtained, and which will be given by (25) or its equivalent (38), can be obtained as follows. Let there be added to the given electrical system a continuous distribution of (generator) constraining voltages such that the approximate result given by (25) without these voltages is the exact solution of the problem with them. These constraining voltages contribute a term $\delta W_s'/\delta q_k$ to (11), and terms $\delta W_s'/\delta q_1, \delta W_s'/\delta q_2, \cdots, \delta W_s'/\delta q_n$ to the right-hand sides of (25), respectively. Denoting the modified equations by (11′) and (25′), respectively, we note that (11) and (25) are approximate for the case of no constraining voltages, whereas (11′) and (25′) are exact for the case where constraining voltages are included. Subtracting each equation of (25) from the corresponding equation of (25′) noting that both systems of equations have the same solution, it follows that

$$(\delta W_s'/\delta q_k) = 0, \quad k = 1, 2, \cdots, n;$$

hence the constraining voltages do no work in any virtual displacement δq_k, $k = 1, 2, \cdots, n$. We have now shown that *the approximate solution given by (25) is the exact solution of the problem obtained by superimposing on the given electrical system a suitable system of constraining voltages which do no work for any virtual displacement* δq_k. We have also obtained the following result. *Let there be added to the given system a set of constraining voltages which forces the solution of the problem to be given exactly (at each instant) by a linear combination of the assumed modes; then (25) gives the one and only such combination that does not require the constraining voltages to do work under any virtual displacement* δq_k *(and hence any virtual displacement permitted by the chosen modes).* It may be noted that the equations (25) become (38), (44), and (51) in the corresponding cases.

4. DERIVATION OF LAGRANGE'S EQUATIONS FOR THE ELECTROMAGNETIC FIELD

Of the three "generalized voltages" appearing in (11) we shall first determine $\delta W_b/\delta q_k$. Noting (10), (2), and the fact that the complete field is sufficiently obtained by superimposing the fields of the N modes, we see that the contribution of any elementary tube of flow to δW_b is

$$\left[10^{-8}\int_a\left(\sum_{i=1}^{N}\dot{B}_i\right)\cdot da\right]$$

$$\times[\text{Virtual charge that flows in tube}]. \quad (12)$$

We now associate with the ith mode an "induced back voltage" function $\bar{E}_i$ which satisfies the relation

$$\nabla\times\bar{E}_i=10^{-8}\dot{B}_i, \quad (13)$$

which is possible since $\dot{B}_i$ is solenoidal. (13) determines $\bar{E}_i$ only to within an arbitrary additive irrotational vector function. Any one of the possible choices of $\bar{E}_i$ will serve our purpose; however, it is convenient to visualize $\bar{E}_i$ as the induced back voltage due to the magnetic field of the ith mode with the boundary surface composed as described in Section 2 except with the surface layer of generators removed. Substituting (13) in the first factor of (12), and applying Stoke's theorem to obtain the line integral

$$\int_s(\sum_{i=1}^{N}\bar{E}_i)\cdot 1_s|ds|,$$

where 1_s is a unit vector along the tube; and writing for the second factor

$$\left|\sum_{i=1}^{N}f_i\frac{\partial Q_i}{\partial q_k}\delta q_k\right|d\sigma,$$

in the volume and

$$\left|\sum_{i=1}^{N}f_{si}\frac{\partial Q_i}{\partial q_k}\delta q_k\right|d\tau,$$

on the bounding surface, respectively, $d\sigma$ being the cross-sectional area of the tube and $d\tau$ being the width of the strip (degenerate tube), (12) becomes the sum of an integral over the volume elements $|ds|d\sigma$ of the tube in space and an integral over the surface elements $|ds|d\tau$ of the strip on the surface. Noting that the direction of the expressions inside the absolute value signs above is that of 1_s, and summing over all the tubes we obtain

$$\delta W_b=\int_V\left[\sum_{i=1}^{N}\bar{E}_i\right]\cdot\left[\sum_{i=1}^{N}f_i\frac{\partial Q_i}{\partial q_k}\right]\delta q_k dV$$

$$+\int_A\left[\sum_{i=1}^{N}\bar{E}_i\right]\cdot\left[\sum_{i=1}^{N}f_{si}\frac{\partial Q_i}{\partial q_k}\right]\delta q_k|dA|,$$

where dA is the vector element of area. Dividing by δq_k, multiplying out the sums in the integrands, and inverting the order of summation and integration, this becomes

$$\frac{\delta W_b}{\delta q_k}=\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial Q_j}{\partial q_k}\left\{\int_V\bar{E}_i\cdot f_j dV+\int_A\bar{E}_i\cdot f_{sj}|dA|\right\},$$

which, noting (6) and (8), can be written

$$\frac{\delta W_b}{\delta q_k}=\frac{10}{4\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial Q_j}{\partial q_k}\left\{\int_V\bar{E}_i\cdot\nabla\times\left(\frac{g_j}{\mu}\right)dV\right.$$

$$\left.+\int_A\bar{E}_i\cdot\left(\frac{g_j}{\mu}\right)\times 1_n|dA|\right\}. \quad (14)$$

Using the vector relations

$$A\cdot\nabla\times B=B\cdot\nabla\times A-\nabla\cdot A\times B,$$

and

$$A\cdot B\times C=C\cdot A\times B, \quad (15)$$

it follows from (13) that (14) can be written

$$\frac{\delta W_b}{\delta q_k}=\frac{10}{4\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial Q_j}{\partial q_k}$$

$$\times\left\{\int_V\left[\left(\frac{g_j}{\mu}\right)\cdot(10^{-8}\dot{B}_i)-\nabla\cdot\bar{E}_i\times\left(\frac{g_j}{\mu}\right)\right]dV\right.$$

$$\left.+\int_A\bar{E}_i\times\left(\frac{g_j}{\mu}\right)\cdot dA\right\},$$

which, applying Gauss' theorem and noting (5), reduces to

$$\frac{\delta W_b}{\delta q_k}=\frac{10^{-7}}{4\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial Q_j}{\partial q_k}\ddot{Q}_i\int_V\frac{g_i\cdot g_j}{\mu}dV. \quad (16)$$

But

$$\ddot{Q}_i\frac{\partial Q_j}{\partial q_k}=\frac{d}{dt}\left(\dot{Q}_i\frac{\partial Q_j}{\partial q_k}\right)-\dot{Q}_i\frac{d}{dt}\left(\frac{\partial Q_j}{\partial q_k}\right); \quad (17)$$

also noting (9) and the fact that

$$\dot{Q}_i=\frac{\partial Q_i}{\partial t}+\sum_{r=1}^{n}\frac{\partial Q_i}{\partial q_r}\dot{q}_r, \tag{18}$$

it can easily be verified by direct differentiation that

$$\frac{\partial Q_j}{\partial q_k}=\frac{\partial \dot{Q}_j}{\partial \dot{q}_k} \quad \text{and} \quad \frac{d}{dt}\left(\frac{\partial Q_j}{\partial q_k}\right)=\frac{\partial \dot{Q}_j}{\partial q_k}, \tag{19}$$

in which expressions Q_j is considered to be a function of $2n+1$ independent variables: the q's, $\dot{q}$'s, and t. Substituting (19) in (17), and then (17) in (16), we obtain

$$\frac{\delta W_b}{\delta q_k}=\frac{10^{-7}}{4\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}G_{ij}\left[\frac{d}{dt}\left(\dot{Q}_i\frac{\partial \dot{Q}_j}{\partial \dot{q}_k}\right)-\dot{Q}_i\frac{\partial \dot{Q}_j}{\partial q_k}\right], \tag{20}$$

where

$$G_{ij}=G_{ji}=\int_V \frac{g_i\cdot g_j}{\mu}dV. \tag{21}$$

The magnetic energy is, noting (5),

$$T=\frac{10^{-7}}{8\pi}\int_V \frac{1}{\mu}\left[\sum_{i=1}^{N}\dot{Q}_i g_i\right]^2 dV \text{ joules}. \tag{22}$$

Squaring out the integrand, inverting the order of summation and integration, and noting (21), this becomes

$$T=\frac{10^{-7}}{8\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}G_{ij}\dot{Q}_i\dot{Q}_j \text{ joules}. \tag{23}$$

Differentiating, noting that $G_{ij}=G_{ji}$, we obtain the relations

$$\frac{\partial T}{\partial \dot{q}_k}=\frac{10^{-7}}{4\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}G_{ij}\dot{Q}_i\frac{\partial \dot{Q}_j}{\partial \dot{q}_k},$$

$$\frac{\partial T}{\partial q_k}=\frac{10^{-7}}{4\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}G_{ij}\dot{Q}_i\frac{\partial \dot{Q}_j}{\partial q_k},$$

which in (20) give

$$\frac{\delta W_b}{\delta q_k}=\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right)-\frac{\partial T}{\partial q_k}. \tag{24}$$

Substituting in (11) we obtain the equations

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right)-\frac{\partial T}{\partial q_k}=\frac{\delta W_g}{\delta q_k}-\frac{\delta W_a}{\delta q_k}, \quad k=1, 2, \cdots, n, \tag{25}$$

which are formally the same as Lagrange's equations in mechanics.[6]

5. CALCULATION OF THE GENERALIZED VOLTAGES

A. Generalized Generator Voltage

Let the e.m.f. per cm produced by a volume distribution of generators be E_g; then the e.m.f. per cm of the surface distributions on the boundary surfaces is the tangential component of E_g-E. (See Section 2.)[7] Noting (2), (4), (5), and (7) it follows that

$$\delta W_a=\int_V E_g\cdot\left(\sum_{i=1}^{N}\frac{\partial Q_i}{\partial q_k}f_i\delta q_k\right)dV$$

$$+\int_A\left(E_g-\frac{4\pi c^2 10^{-9}}{K}\sum_{i=1}^{N}Q_i f_i\right)$$

$$\cdot\left(\frac{10}{4\pi}\sum_{i=1}^{N}\frac{\partial Q_i}{\partial q_k}\left(\frac{g_i}{\mu}\right)\times 1_n\right)\delta q_k|dA|.$$

The second factor in each integral is a virtual charge density (charge being time integral of current). We note that the surface charge per cm is given by replacing $\dot{Q}_i$ by $\delta Q_i=(\partial Q_i/\partial q_k)\delta q_k$ in (5) and (7). Applying the second relation of (15) we obtain finally

$$\frac{\delta W_a}{\delta q_k}=\sum_{i=1}^{N}\mathcal{E}_i\frac{\partial Q_i}{\partial q_k}, \tag{26}$$

where

$$\mathcal{E}_i=\int_V E_g\cdot f_i dV+\frac{10}{4\pi}\int_A E_g\times\left(\frac{g_i}{\mu}\right)\cdot dA$$

$$-10^{-8}c^2\sum_{j=1}^{N}Q_j\int_A\frac{f_j\times g_i}{K\mu}\cdot dA. \tag{27}$$

The first two of these integrals vanish if there is no volume distribution of generators.

B. Generalized Inductance Drop

We see from (24) that the generalized inductance drop $\delta W_b/\delta q_k$ forms the left hand side of Lagrange's equations (25).

[7] If the volume distribution of generators extends to the boundary surface, E must be replaced by $E-E_g$ in (1).

C. Generalized Capacitance Drop

The virtual work done on capacitance drops (voltages arising from the flow of displacement currents) is

$$\delta W_{ac}=4\pi c^2 10^{-9}\int_V\left(\frac{1}{K}\sum_{i=1}^{N}Q_i f_i\right)\cdot\left(\sum_{i=1}^{N}\frac{\partial Q_i}{\partial q_k}f_i\delta q_k\right)dV,$$

the first factor being the capacitance drop per cm, and the second being the virtual charge per cm² that flows. Multiplying out the integrand and inverting the order of summation and integration, we obtain the generalized capacitance drop

$$\frac{\delta W_{ac}}{\delta q_k}=4\pi c^2 10^{-9}\sum_{i=1}^{N}\sum_{j=1}^{N}F_{ij}Q_i\frac{\partial Q_j}{\partial q_k}, \tag{28}$$

which is one part of $\delta W_a/\delta q_k$. Here

$$F_{ij}=F_{ji}=\int_V\frac{1}{K}(f_i\cdot f_j)dV, \tag{29}$$

the integration extending over those regions where displacement currents flow.

The electric energy, given by an integral similar to (22), is

$$\Phi=2\pi c^2 10^{-9}\sum_{i=1}^{N}\sum_{j=1}^{N}F_{ij}Q_iQ_j; \tag{30}$$

hence, noting the symmetry of the F's we see that the generalized capacitance drop is

$$\frac{\delta W_{ac}}{\delta q_k}=\frac{\partial\Phi}{\partial q_k}, \tag{31}$$

as was to be expected.

D. Generalized Resistance Drop

The virtual work done on resistance drops is

$$\delta W_{ar}=\int_V(\rho\sum_{i=1}^{N}\dot{Q}_i f_i)\cdot\left(\sum_{i=1}^{N}\frac{\partial Q_i}{\partial q_k}f_i\delta q_k\right)dV, \tag{32}$$

the first factor in the integrand being the resistance drop per cm caused by the resistivity $\rho(x, y, z)$ ohm cm, and the second being the charge per cm² that flows. Multiplying out the integrand, and inverting the order of summation and integration, we obtain for the contribution to $\delta W_a/\delta q_k$ the "generalized resistance drop"

$$\frac{\delta W_{ar}}{\delta q_k}=\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}\dot{Q}_i\frac{\partial Q_j}{\partial q_k}, \tag{33}$$

where

$$R_{ij}=R_{ji}=\int_V\rho(f_i\cdot f_j)dV, \tag{34}$$

the volume integral extending over those regions where the currents must pass through resistive material. The power dissipated at any instant is

$$P=\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}\dot{Q}_i\dot{Q}_j, \tag{35}$$

which is given by an integral similar to (32). Noting (19) and the symmetry of the R's, we see that

$$\frac{\delta W_{ar}}{\delta q_k}=\frac{1}{2}\frac{\partial P}{\partial\dot{q}_k}. \tag{36}$$

It should be noted that the dissipation just considered is "series" dissipation in the sense that the specified mode currents must pass through resistive material. At first sight it may seem that the results obtained are not applicable in cases where the dissipation is "shunt" in the sense that because of the resistivity an alternative path is provided for the mode currents, as, for example, in the case of the leakage conductance of a dielectric. This difficulty disappears, however, if one keeps in mind the following two facts.

a. Once a current mode has been chosen, the geometric pattern of the currents of that mode is fixed—there is no such thing as "an alternative path to a mode current."

b. The number and natures of the current modes must be such that a sufficient approximation to the actual current distribution is possible.

In the case of an electromagnetic field containing a leaky dielectric, for example, the current modes may first be chosen in the same manner as if there were no leakage.[8] For each of

[8] This is true only if the conduction currents are so small in comparison with the displacement currents that the leakage does not appreciably alter the shape of the electric field. If the leakage is large enough to appreciably alter this shape, enough suitable modes must be chosen to permit the field to be approximated at each instant by a linear combination of these modes.

the modes

$$J_i=\dot{Q}_i f_i \text{ amp./cm}^2,$$

which involves displacement current in the dielectric, a companion mode

$$\bar{J}_i=\bar{Q}_i'\bar{f}_i \text{ amp./cm}^2,$$

is now chosen as follows.

$$\begin{aligned}&\bar{f}_i=0 \text{ outside of the dielectric,}\\ &\bar{f}_i=f_i \text{ in the dielectric}\\ &\text{(conduction current density vector)}\\ &=-\text{(displacement current density vector)}\\ &=\bar{Q}_i'\bar{f}_i \text{ amps./cm}^2.\end{aligned}\tag{37}$$

Here a prime, like a dot, indicates differentiation with respect to time. This companion mode evidently has the following rather peculiar properties.

a. The conduction current and displacement current vanish at all points outside of the dielectric, and cancel each other at all points in the dielectric; hence the total current is everywhere zero, and no magnetic field is produced.

b. In the dielectric the conduction current and the displacement current distributions are geometrically similar to those of the first chosen mode J_i. For values of $\dot{Q}_i$ and $\bar{Q}_i'$ which are of like sign, the displacement currents of $\bar{J}_i$ oppose those of J_i, whereas the conduction currents of $\bar{J}_i$ flow in the same direction as the displacement currents of J_i.

c. The net effect of the companion mode is to provide in a sense the alternative path indicated by the presence of the leakage. It should be noted that this path requires the addition of a new mode—not a change in the choice of the original mode J_i.[8]

Substituting (26), (31), and (36) in (25), we obtain

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right)-\frac{\partial T}{\partial q_k}+\frac{\partial \Phi}{\partial q_k}+\frac{1}{2}\frac{\partial P}{\partial \dot{q}_k}=\sum_{i=1}^{N}\mathcal{E}_i\frac{\partial Q_i}{\partial q_k},\quad k=1, 2, \cdots, n,\tag{38}$$

$\mathcal{E}_i$ being given by (27). This is the desired form of Lagrange's equations.[6,9]

[9] In the special case of a closed cavity with a perfectly conducting boundary surface Hamilton's principle for the electromagnetic field follows from (38), and vice versa (see reference 1).

6. SPECIAL CASE WHERE $Q_i=q_i$; EQUIVALENT NETWORKS

We shall now consider the special case where

$$Q_i=q_i,\quad i=1, 2, \cdots, n,\tag{39}$$

which is sufficient to cover all work that has been done to date on the echo box. In this case

$$\frac{\partial Q_i}{\partial t}=0,\quad \frac{\partial Q_i}{\partial q_r}=\begin{cases}1 & \text{if } i=r\\ 0 & \text{if } i\neq r\end{cases};$$

and (23), (30), and (35) become

$$\begin{aligned}T&=\frac{10^{-7}}{8\pi}\sum_{i=1}^{n}\sum_{j=1}^{n}G_{ij}\dot{q}_i\dot{q}_j \text{ joules},\\ \Phi&=2\pi c^2 10^{-9}\sum_{i=1}^{n}\sum_{j=1}^{n}F_{ij}q_iq_j \text{ joules},\\ P&=\sum_{i=1}^{n}\sum_{j=1}^{n}R_{ij}\dot{q}_i\dot{q}_j \text{ watts},\end{aligned}\tag{40}$$

where

$$\begin{aligned}F_{ij}&=F_{ji}=\int_V\frac{1}{K}(f_i\cdot f_j)dV,\\ G_{ij}&=G_{ji}=\int_V\frac{1}{\mu}(g_i\cdot g_j)dV,\\ R_{ij}&=R_{ji}=\int_V\rho(f_i\cdot f_j)dV.\end{aligned}\tag{41}$$

Also, (27) becomes

$$\mathcal{E}_i=\int_V E_g\cdot f_i dV+\frac{10}{4\pi}\int_A E_g\times\left(\frac{g_i}{\mu}\right)\cdot dA-10^{-8}c^2\sum_{j=1}^{n}q_j\int_A\frac{f_j\times g_i}{K\mu}\cdot dA.\tag{42}$$

Substituting these expressions in (38) Lagrange's equations become[10]

$$\sum_{i=1}^{n}\left[\frac{10^{-7}}{4\pi}G_{ki}\frac{d^2}{dt^2}+R_{ki}\frac{d}{dt}+4\pi c^2 10^{-9}F_{ki}\right]q_i=\mathcal{E}_k,\quad k=1, 2, \cdots, n.\tag{43}$$

[10] The first terms in the brackets of (43) and (50), which give the generalized inductance drops, can be obtained directly from (16), it being therefore unnecessary to transform (16) into the Lagrangean form (24) for this purpose. By obtaining Lagrange's equations, however, nothing is lost, and the theory is rendered more complete.

These equations are identical in form with those obtained by applying Kirchoff's laws to a lumped network in which the q's are mesh charges, which fact is more apparent if (43) is written

$$\sum_{i=1}^{n} Z_{ki} I_i = \mathcal{E}_k, \quad k=1, 2, \cdots, n, \tag{44}$$

where

$$I_i = \dot{q}_i, \qquad Z_{ij} = L_{ij}p + R_{ij} + \frac{1}{pC_{ij}},$$

$$L_{ij} = \frac{10^{-7}}{4\pi} G_{ij}, \quad C_{ij} = \frac{1}{4\pi c^2 10^{-9} F_{ij}}, \tag{45}$$

$$p = \frac{d}{dt}, \qquad \frac{1}{p} = \int_{t_0(q's=0)}^{t} dt,$$

which is the notation commonly used in circuit theory.[6] We thus see that Lagrange's equations are in a sense a generalization of Kirchoff's laws, the $\mathcal{E}$'s, I's, Z's, L's, R's, and C's of (44) and (45) being "generalized voltages, currents, impedances, inductances, resistances, and capacitances," respectively. Furthermore we see from (41) and (45) that the impedance components are symmetrical, thus

$$L_{ij} = L_{ji}, \quad R_{ij} = R_{ji}, \quad C_{ij} = C_{ji}. \tag{46}$$

Since T, Φ, and P, being energies and power loss, are intrinsically positive quantities, it follows from (40) and (45) that the quadratic forms

$$\sum_{i=1}^{n}\sum_{j=1}^{n} L_{ij}x_i x_j, \quad \sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij}x_i x_j, \quad \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{C_{ij}} x_i x_j,$$

are all positive definite; hence a lumped network exists whose mesh currents and voltages are the I's and $\mathcal{E}$'s, whose impedance components are the L's, R's, and C's, and whose mesh equations obtained from Kirchoff's laws duplicate (44). It may, however, contain ideal transformers. *This network is equivalent to the given electromagnetic system insofar as is permitted by the choice of the number and kind of modes.*[3,6] It is evident that once the Eqs. (44) have been obtained, or, what is the same thing, the matrices or tables of values of the L's, R's, C's, and $\mathcal{E}$'s, these matrices can be used to devise specific equivalent networks. If the solution of the equations (44) is to be entirely mathematical, there is no particular object in looking for an equivalent network, since such a network is by definition one whose mesh equations duplicate (44). If, however, the solution is to be obtained using a network analyzer (phantom circuit or "model"), then an equivalent network must be devised *which can be realized physically*. Aside from such a use, however, an equivalent network gives one an intuitive feeling for the significance of the various modes and design constants that is not given by the matrices composed of the Z's and $\mathcal{E}$'s of (44); also it enables one to estimate the behavior of the given electromagnetic system as a component part of a larger network or system, and hence the behavior of the system as a whole. Because of this the equivalent network is useful in guiding certain types of experimental work.

7. SPECIAL CASE WHERE EACH Q IS A CONSTANT LINEAR COMBINATION OF THE q's

We shall next consider the special case where

$$Q_i = \sum_{j=1}^{n} a_{ij} q_j, \quad i=1, 2, \cdots, N, \tag{47}$$

the a's being constants; then

$$\frac{\partial Q}{\partial t} = 0, \quad \frac{\partial Q_i}{\partial \dot{q}_r} = a_{ir};$$

and (23), (30), and (35) become

$$T = \frac{10^{-7}}{8\pi} \sum_{r=1}^{n}\sum_{s=1}^{n} \dot{q}_r \dot{q}_s \sum_{i=1}^{N}\sum_{j=1}^{N} G_{ij} a_{ir} a_{js} \text{ joules},$$

$$\Phi = 2\pi c^2 10^{-9} \sum_{r=1}^{n}\sum_{s=1}^{n} q_r q_s \sum_{i=1}^{N}\sum_{j=1}^{N} F_{ij} a_{ir} a_{js} \text{ joules}, \tag{48}$$

$$P = \sum_{r=1}^{n}\sum_{s=1}^{n} \dot{q}_r \dot{q}_s \sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij} a_{ir} a_{js} \text{ watts},$$

respectively, where the F's, G's, and R's are again given by (41). Denoting the coefficient on the right of $\dot{q}_r\dot{q}_s$ in T, of $q_r q_s$ in Φ, and of $\dot{q}_r\dot{q}_s$ in P by T_{rs}, Φ_{rs}, and P_{rs}, respectively, it follows from (41) and the form of these quantities in (48) that

$$\begin{aligned} T_{rs} &= T_{sr} = \|a_{ir}\| \cdot \|G_{ij}\| \cdot \|a_{js}\|, \\ \Phi_{rs} &= \Phi_{sr} = \|a_{ir}\| \cdot \|F_{ij}\| \cdot \|a_{js}\|, \\ P_{rs} &= P_{sr} = \|a_{ir}\| \cdot \|R_{ij}\| \cdot \|a_{js}\|, \end{aligned} \tag{49}$$

where in each of these expressions the row indices of the first, second, and third matrix factors are r, i, and j, respectively. The multiplications are the usual row times column matrix multiplications, which are associative but not commutative.

Substituting these expressions in (38), Lagrange's equations become[10]

$$\sum_{i=1}^{n}\left[\frac{10^{-7}}{4\pi}T_{ki}\frac{d^2}{dt^2}+P_{ki}\frac{d}{dt}+4\pi c^2 10^{-9}\Phi_{ki}\right]q_i=\bar{\mathcal{E}}_k,$$

$$k=1, 2, \cdots, n, \quad (50)$$

where

$$\bar{\mathcal{E}}_k=\sum_{i=1}^{N}\mathcal{E}_i a_{ik},$$

$\mathcal{E}_i$ being given by (27). Proceeding as with (43) this may be written[6]

$$\sum_{i=1}^{n}Z_{ki}I_i=\bar{\mathcal{E}}_k, \quad k=1,2, \quad \cdots, n, \quad (51)$$

where

$$I_i=\dot{q}_i,$$

$$Z_{ij}=L_{ij}p+\mathcal{R}_{ij}+\frac{1}{pC_{ij}}$$

$$L_{ij}=\frac{10^{-7}}{4\pi}T_{ij}, \quad \mathcal{R}_{ij}=P_{ij}, \quad C_{ij}=\frac{1}{4\pi c^2 10^{-9}\Phi_{ij}}, \quad (52)$$

$$p=\frac{d}{dt}, \qquad \frac{1}{p}=\int_{h(q's=0)}^{t}dt.$$

Noting (49) and (52) we see that

$$L_{ij}=L_{ji}, \quad \mathcal{R}_{ij}=\mathcal{R}_{ji}, \quad C_{ij}=C_{ji}, \quad Z_{ij}=Z_{ji}; \quad (53)$$

also since T, Φ, and P are intrinsically positive, it follows from (48) that the quadratic forms

$$\sum_{i=1}^{n}\sum_{j=1}^{n}L_{ij}x_i x_j, \quad \sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{R}_{ij}x_i x_j, \quad \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{C_{ij}}x_i x_j,$$

are positive definite; *hence a lumped network exists whose mesh equations are (51), and which is therefore equivalent to the given electromagnetic system in so far as is permitted by the choice of the number and kind of modes.*[3,6] This network may, however, contain ideal transformers. We thus see that Lagrange's equations again give a generalization of Kirchoff's laws. The remarks made at the end of Section 6 concerning equivalent networks also apply here.

8. UNITS

Where it has been necessary to specify units, we have so far used the following: centimeters, seconds, volts, coulombs, amperes, joules. It has not as yet been necessary to assign units to the q's. In Sections 6 and 7, however, where the special cases $Q_i=q_i$ and

$$Q_i=\sum_{r=1}^{n}a_{ir}q_r,$$

were considered, it was shown that a lumped network exists such that if the charges that flow in the various meshes are the q's, then the mesh equations for the network duplicate Lagrange's equations for the given electromagnetic system. It is hence advantageous in these two cases to assign to the q's the units of charge, namely coulombs. This we shall do; also we shall let the a_{ij}'s be pure numbers, so that the Q_i's are also expressed in coulombs. It follows immediately that f_i is the fraction of the charge Q_i that flows per cm^2, f_{si} the fraction that flows per cm on the surface; and that each term of Lagrange's equations is a voltage, since it is the quotient of an energy variation and a charge variation.[3,5] It also follows that L_{ij} is in henries, R_{ij} and Z_{ij} are in ohms, C_{ij} is in farads, and $\mathcal{E}_k$ and $\bar{\mathcal{E}}_k$ are in volts.[11] It may be mentioned that c = velocity of light $=2.99796\cdot 10^{10}$ cm/sec.

Let us suppose that the given electromagnetic system contains a pair of terminals at which it connects with an external electrical system, for example the ends of a coupling loop in the case of an echo box; then if the charge passing through either terminal be taken as one of the q's, say q_0, it follows that the resulting equivalent network will have a mesh which has a current $\dot{q}_0$ and an external source voltage $\mathcal{E}_0$ or $\bar{\mathcal{E}}_0$ equal to the terminal current and voltage, respectively, in the actual system. We thus see that the external electrical system would not be affected if the given electromagnetic system were replaced by

[11] It is evidently possible to devise an equivalent network, the dual of that considered above, in which the q's are node voltages instead of mesh charges, and for which the various terms in Lagrange's equations are generalized currents instead of generalized voltages. If, however, this network is to replace the given electromagnetic system in a larger network, it is necessary that this larger network be replaced by its dual. It appears more natural to associate current modes with mesh currents, as is done above, than to associate them with node voltages.

its equivalent network; and furthermore that this network would receive the proper excitation, or source voltage $\mathcal{E}_0$ or $\bar{\mathcal{E}}_0$ from the external system. A similar situation exists if there are M terminals instead of two common to the given system and the external system, for the currents passing through these M terminals can be represented by $(M-1)$ circulating currents each of which flows in one terminal and out another, and each of these circulating currents can be taken as a $\dot{q}$.

9. CASES WHERE $Q_i \neq q_i$; MODES WHICH CHANGE WITH TIME; MOVING COORDINATE SYSTEMS

Let us choose a set of n modes which are capable of representing the electromagnetic field with sufficient accuracy, the corresponding f's, g's, and Q's being denoted by $\bar{f}$'s, $\bar{g}$'s, and q's, respectively, the q's being taken as generalized coordinates. Now let us suppose that a very large or infinite number N of modes have been chosen which are capable of representing the actual electromagnetic field and also each of the above n modes so closely that the differences are undetectable by any physical means, the corresponding f's, g's, and Q's being denoted by f's, g's, and Q's, respectively. For example the f's might be a complete set of orthogonal vector functions which satisfy the required boundary conditions. We now have

$$\bar{f}_i = \sum_{j=1}^{N} A_{ij} f_j, \quad i = 1, 2, \cdots, n, \tag{54}$$

since each of the n modes can be represented by the large set, the A's being the corresponding coefficients in the expansions. From (54) it follows that

$$\sum_{i=1}^{n} q_i \bar{f}_i = \sum_{i=1}^{n} q_i \sum_{j=1}^{N} A_{ij} f_j = \sum_{j=1}^{N} f_j \sum_{i=1}^{n} q_i A_{ij}.$$

Placing

$$Q_j = \sum_{i=1}^{n} q_i A_{ij}, \tag{55}$$

it follows that

$$\sum_{i=1}^{n} q_i \bar{f}_i = \sum_{j=1}^{N} Q_j f_j.$$

From this we see that the approximation obtained by using the n modes directly with $Q_i = q_i$ as in Section 6 can also be obtained by using the large set of N modes and choosing (55) as the relation (9), which gives the Q's in terms of the q's and t.

If the A's are constants we obtain a case of the type treated in Section 7, and hence the same results, equivalent networks for example, that would be obtained using the smaller set of n modes and the procedures of Section 6.

Since t appears in (9), Lagrange's equations are valid if the A's in (55) are any specified functions of time. From (54) we see that specifying the A's as functions of time is equivalent to choosing the f's of the small set of n modes as functions of position *and time*, thus

$$\bar{f}_i = \bar{f}_i(x, y, z, t).$$

The electric charge per cm² at any point is

$$\sum_{i=1}^{n} q_i \bar{f}_i(x, y, z, t);$$

hence we have obtained the following result. *If a set of changing modes (f's functions of time as well as position) is chosen which is capable of approximating the actual field with sufficient accuracy, and if the magnification factors associated with the various modes are taken as generalized coordinates, then Lagrange's equations are valid.* The current density vector in such a case for any mode is

$$J_i = \dot{q}_i \bar{f}_i + q_i \bar{f}_i' \text{ amps./cm}^2.$$

Since H_i is determined by the relation

$$\nabla \times H_i = 0.4\pi J_i,$$

it follows that

$$B_i = \dot{q}_i \bar{g}_i + q_i \bar{g}_i' \text{ gauss}, \tag{56}$$

where

$$\nabla \times \left(\frac{\bar{g}_i}{\mu} \right) = 0.4\pi \bar{f}_i.$$

We see that (6) is still valid, but that (56) replaces (5) in computing T.

As an application of changing modes we see that Lagrange's equations are valid if a set of modes are used which are seen as unchanging by an observer moving in any specified manner, for a stationary observer would see these same

TABLE I.

Problem considered	Procedure using electric current modes	Procedure using magnetic flux modes
Coupling loop	Requires extra mode.	Treated as magnetic generator.
Series dissipation	Gives generalized resistance drop.	Requires companion modes. (See Eq. (37).)
Shunt dissipation	Requires companion modes.	Gives generalized m.m.f.

modes as changing. As a further application it appears that changing modes might be useful in some wave propagation problems.

In certain problems it may be convenient to introduce one or more equations of constraint involving the q's and t. These can be handled by Lagrange's method of multipliers, as in mechanics.

10. DERIVATION OF LAGRANGE'S EQUATIONS USING MAGNETIC FLUX MODES

There is a duality between electric quantities and magnetic quantities which permits these two types of quantities to exchange their roles in the application of Lagrangian procedures. This is not evident if the practical system of units is used (volts, amp., etc.); hence *in this one section* we shall use the Gaussian system of units, in which electrical quantities are expressed in absolute electrostatic units, and magnetic quantities are expressed in absolute electromagnetic units; and note that these two systems of units are dual in that a magnetic charge is used to build up one system in exactly the same manner in which an electric charge is used to build up the other. To complete the duality we must introduce the concept of magnetic current, this being the rate of change of the magnetic charge associated with the corresponding flux. The surface and volume magnetic current densities are therefore

$$\mathcal{J}_s=(\dot{B}_s/4\pi)\text{ abs. e.m.u./cm},$$

and

$$\mathcal{J}=(\dot{B}/4\pi)\text{ abs. e.m.u./cm}^2, \qquad (57)$$

respectively.

Maxwell's equations for free space, and the equations which determine the surface electric and magnetic current distributions which can bound an electromagnetic field are

$$\begin{aligned} &\nabla\cdot B=0, && \nabla\cdot D=0,\\ &\nabla\times H=\frac{4\pi J}{c}, && \nabla\times E=-\frac{\dot{B}}{c}=-\frac{4\pi\mathcal{J}}{c},\\ &B=\mu H, && D=KE,\\ &\mathcal{J}_s=\frac{\dot{B}_s}{4\pi}=-\frac{c}{4\pi}E\times 1_n, && J_s=\frac{c}{4\pi}H\times 1_n. \end{aligned} \qquad (58)$$

If these equations are rewritten using the following notation:

$$\begin{aligned} &E=\text{magnetic field intensity},\\ &D=\text{magnetic flux density},\\ &J=\text{magnetic volume current density},\\ &J_s=\text{magnetic surface current density},\\ &H=-(\text{electric field intensity}),\\ &B=-(\text{electric flux density}),\\ &\mathcal{J}=-\text{electric volume current density},\\ &\mathcal{J}_s=-\text{electric surface current density},\\ &K=\text{permeability},\\ &\mu=\text{dielectric constant}; \end{aligned} \qquad (59)$$

it is evident that the resulting equations duplicate (58), although they will be obtained in a different order. We now see that the mathematical work in the preceding sections applies to establish the Lagrangian procedures associated with the notation (59).[12] In these the roles of the electric and magnetic quantities are interchanged, and we note that T is the electric energy, Φ is the magnetic energy, E_g is a volume distribution of m.m.f.; also that on those portions of the bounding surface which are not perfect electric or magnetic conductors the surface magnetic currents lie just *inside* and completely shield the surface electric currents, and pass through a surface distribution of m.m.f. In the virtual displacement used in calculating the generalized m.m.f.'s the m.m.f.'s existing at time t are considered as persisting unchanged.

Whether it is better to use electric current modes or magnetic flux modes in working any specific problem depends entirely upon the nature of the problem. As might be expected the

[12] The units used in this work would bear the same relation to magnetic quantities as they did to electric quantities before, thus 1 coulomb of magnetic charge is the magnetic charge associated with a flux of $0.4\pi c$ maxwells, and 1 volt of magnetic potential difference is $10^8/c$ oersteds, etc.

procedures for handling a particular phenomenon are different in the two cases.

Certain of these differences are indicated in Table I.

APPENDIX

The Bounding of an Electromagnetic Field by Surface Electric and Magnetic Currents

Using Gaussian units, denoting the unit vector normal to the bounding surface outward by 1_n, and defining magnetic current by (57), it is evident that the field inside the surface can be bounded so that there is no external field if there are surface electric and magnetic currents of surface densities

$$J_s=\frac{c}{4\pi}H\times 1_n, \qquad \mathfrak{J}_s=-\frac{c}{4\pi}E\times 1_n. \tag{1}$$

We must, however, show that the surface divergence of J_s and $\mathfrak{J}_s$ are equal to the normal components of $\dot{D}/4\pi$ and $\dot{B}/4\pi$, respectively, so that there will be no discontinuity of electric or magnetic current.

Let us choose any closed contour C in the surface, then the total electric current leaving this contour is, noting (1),

$$\int_C 1_n\times J_s\cdot ds=\int_C 1_n\times\left[\frac{c}{4\pi}H\times 1_n\right]\cdot ds, \tag{2}$$

where ds is the vector element of arc length. Here the cross multiplication merely rotates the surface current vector through 90° so that the tangential component of the product vector is the normal component of J_s. For continuity (2) must equal

$$\int_A\frac{\dot{D}}{4\pi}\cdot dA, \tag{3}$$

where A is the area enclosed by C and dA is the vector element of area. Applying Stoke's theorem to (2) and comparing the result with (3) noting that C is arbitrary, we see that our problem is to show that[13]

$$1_n\cdot\nabla\times\left\{1_n\times\left[\frac{c}{4\pi}H\times 1_n\right]\right\}=\frac{\dot{D}}{4\pi}\cdot 1_n. \tag{4}$$

Because of the vector relation

$$A\times(B\times C)=(A\cdot C)B-(A\cdot B)C,$$

and the Maxwell equation

$$\nabla\times H=\frac{\dot{D}}{c}, \tag{5}$$

(4) becomes

$$1_n\cdot\nabla\times[H-(1_n\cdot H)1_n]=(\nabla\times H)\cdot 1_n,$$

or

$$1_n\cdot\nabla\times\mathbf{H}_n=0,$$

where $\mathbf{H}_n$ is a normal vector of magnitude equal to the normal component of H. But

$$1_n\cdot\nabla\times\mathbf{H}_n=\operatorname*{Lim}_{a\to 0}\frac{1}{a}\int_a\mathbf{H}_n\cdot ds=0,$$

due to Stoke's theorem, the integration being over a contour bounding an area a of the surface; hence the desired relation is proved.[13]

Similar reasoning using the second equation of (1) and the Maxwell equation

$$\nabla\times E=-\dot{B}/c$$

instead of (5) can be used to establish the continuity of the magnetic currents. The electric and magnetic currents (1) therefore completely bound the field, leaving it unaltered within the surface and zero outside of the surface, and serving to close all tubes of electric and magnetic current.

[13] Since 1_n is defined only on the surface we see that only the normal component of curl and only integrations which lie in the surface are defined here; but Stoke's theorem applies to these.

An Observation on Heat Sources Placed at Random on a Surface

GREGORY H. WANNIER
Socony-Vacuum Laboratories, Paulsboro, New Jersey
(Received April 26, 1948)

A mean value theorem and a fluctuation theorem are derived for the case where a potential is due to sources of potential placed irregularly on a surface. The purpose of the theorems is to provide a check on theories of solid-solid friction where extremely high "local temperatures" are sometimes assumed. The result indicates that a certain caution should be exercised in using such concepts.

I. BASIC APPROACH

IF sources of potential are placed on a surface, it is not even approximately possible to get a value of the potential by a study of local conditions alone. A finite value for the potential results only if the surface is seen as being embedded in three dimensional space and dissipation of the flux in three dimensions is properly considered.

There are some cases, however, where the problem can be split up in two parts. In the first, the potential is evaluated for the actual geometry with a smooth source function. In the second, local irregularities are introduced as a very erratic source function which averages out to zero in a very small area. The potential due to this second problem is now dependent on local conditions only and the geometry can be drastically simplified.

The purpose of this article is the study of this microscopic effect. We assume then that, within a larger problem of type 1, we can pick a small area satisfying the following conditions:

(a) The area is so small it can be considered plane.

(b) The "smooth" variation of the source function $S(x, y)$ within the area is negligible, so that S may be treated as a constant S_0.

(c) The above statement also applies to the potential function $F(x, y, z)$ whose value will be taken as F_0.

(d) In spite of (a), (b), and (c) the area is still large compared to the irregularities in the source function $S(x, y)$.

With the help of these assumptions, we can write down the basic formula for the potential $F(x, y)$ on the surface (potentials outside the surface will not be considered in the following):

$$F(x, y)-F_0 = \Gamma \int\int_A \frac{S(x', y')-S_0}{((x'-x)^2+(y'-y)^2)^{\frac{1}{2}}} dx'dy', \quad (1)$$

where Γ is some dimensional constant. The formula is worth a few observations. Without the precaution of subtracting S_0 under the integral sign the equation would be quite absurd. Even with this term convergence is by no means obvious. We must postulate a "sufficient irregularity" in S to produce convergence and hence dependence on local conditions only. The integral over A can then be replaced by an integral over an infinite plane.

II. MEAN VALUE THEOREM

If the first moments of the source distribution $S-S_0$ vanish, then the average of the potential $F-F_0$ is also zero.

To prove this we Fourier transform Eq. (1)

$$S(x, y)-S_0 = \int\int e^{i(\mu x+\nu y)}\sigma(\mu, \nu)d\mu d\nu, \quad (2)$$

$$F(x, y)-F_0 = \int\int e^{i(\mu x+\nu y)}\varphi(\mu, \nu)d\mu d\nu. \quad (3)$$

If use is made of the definite integral

$$\int_0^\infty J_0(x)dx = 1,$$

then we get the transformed relation (1)

$$\varphi(\mu, \nu) = \frac{2\pi\Gamma\sigma(\mu, \nu)}{(\mu^2+\nu^2)^{\frac{1}{2}}}. \quad (4)$$

Now, we invert (2) and (3):

$$\sigma(\mu, \nu) = \frac{1}{4\pi^2}\int\int \{S(x, y)-S_0\}e^{-i(\mu x+\nu y)}dxdy,$$

$$\varphi(\mu, \nu) = \frac{1}{4\pi^2}\int\int \{F(x, y)-F_0\}e^{-i(\mu x+\nu y)}dxdy.$$

From this we see that $\sigma(0, 0)$ is 0 by assumption, and that we wish to find the condition under which $\varphi(0, 0)$ vanishes. In view of (4) the conditions are:

$$0=\left(\frac{\partial\sigma}{\partial\mu}\right)_{0,0}=\frac{-i}{4\pi^2}\int\int\{S(x, y)-S_0\}xdxdy,$$

and

$$0=\left(\frac{\partial\sigma}{\partial\nu}\right)_{0,0}=\frac{-i}{4\pi^2}\int\int\{S(x, y)-S_0\}ydxdy.$$

This is the condition on first moments as stated above. The condition implies an "essential" identity of the $+x$ and $-x$ direction. Anisotropy is not excluded, however, that is the $+x$ and $+y$ directions may be different.

Theorem 1 is a preliminary theorem which shows under which conditions F_0 is the correct mean value of F. We come now to a theorem connecting fluctuations.

III. FLUCTUATION THEOREM

The root mean square of the tangential potential gradient is $2\pi\Gamma$ *times the root mean square deviation of the source function.*

The integrals used in the theorem are nonconvergent, of a type which is proportional to the area. This makes the Fourier integral method inconvenient to use. Fourier series are more useful for the present purpose. They are introduced through the well-known device of assuming a very long period $2\pi a$ in both the x and y directions. Equations (2) and (3) are then replaced by

$$S(x, y)-S_0=\sum_{m, n=-\infty}^{+\infty}\Sigma(m, n)e^{i/a(mx+ny)}, \quad (5)$$

$$F(x, y)-F_0=\sum_{m, n=-\infty}^{+\infty}\phi(m, n)e^{i/a(mx+ny)}. \quad (6)$$

The new Fourier coefficients are given in terms of the old transforms by

$$\Sigma(m, n)=\frac{1}{a^2}\sigma\left(\frac{m}{a}, \frac{n}{a}\right),$$

$$\phi(m, n)=\frac{1}{a^2}\varphi\left(\frac{m}{a}, \frac{n}{a}\right),$$

which brings relation (4) into the form:

$$\phi(m, n)=\frac{2\pi\Gamma a\Sigma(m, n)}{(m^2+n^2)^{\frac{1}{2}}}. \quad (7)$$

The mean square deviation of S or F is now simply expressed:

$$\langle(S-S_0)^2\rangle_{Av}=\sum_{m, n=-\infty}^{+\infty}|\Sigma(m, n)|^2, \quad (8)$$

and finally the theorem proved:

$$\left\langle\left(\frac{\partial F}{\partial x}\right)^2+\left(\frac{\partial F}{\partial y}\right)^2\right\rangle_{Av}=\sum_{m n}\frac{m^2+n^2}{a^2}|\phi(m, n)|^2$$
$$=(2\pi\Gamma)^2\sum_{m, n}|\Sigma(m, n)|^2$$
$$=(2\pi\Gamma)^2\langle(S-S_0)^2\rangle_{Av}.$$

IV. AN APPLICATION TO SOLID FRICTION

The fluctuation theorem just proved can be applied in various fields of physics. In electricity, for instance, it would read:

$$\langle E_{||}^2\rangle_{Av}=4\pi^2\langle(\sigma-\sigma_0)^2\rangle_{Av}, \quad (9)$$

where $E_{||}$ is the tangential electric field on the surface of an insulator and σ the actual, σ_0 the "smoothed out" surface charge density.

The author's purpose in developing these theorems was in the field of heat flow. In the theory of friction between two solids, the concept of "point contact" plays a great role. Local temperatures at such point contacts are presumed to be extremely high, giving rise to various high temperature phenomena.

The formula just developed provides a check for such concepts in a given experimental arrangement. It reads:

$$\langle(\nabla T)_{||}^2\rangle_{Av}=\frac{1}{4k^2}\langle(S-S_0)^2\rangle_{Av}. \quad (10)$$

The left-hand side is the mean square of the tangential temperature gradient on the surface of contact, k is the heat conductivity of the two solids which are assumed to be of the same material. S is the actual, S_0 the "smoothed out" heat developed per unit area and unit time.

If we deal with a situation which is essentially isotropic, we can simplify (10) to give the mean square derivative of the temperature in an arbi-

trary tangential direction:

$$\left\langle\left(\frac{dT}{dl}\right)^2\right\rangle_{Av}=\frac{1}{8k^2}\langle(S-S_0)^2\rangle_{Av}. \qquad (11)$$

It is unfortunate that there does not seem to exist a fluctuation theorem involving temperature fluctuations directly since this is usually the quantity of main experimental interest. In order to apply (10) or (11) for this purpose we have to introduce two reasonable relations of a qualitative type. The first one is

$$[\langle(T-T_0)^2\rangle_{Av}]^{\frac{1}{2}}\sim\tfrac{1}{4}[\langle(dT/dl)^2\rangle_{Av}]^{\frac{1}{2}}P. \qquad (12)$$

$\frac{1}{2}P$ is a length which measures the mean distance within which the temperature gradient keeps its sign. P is then the mean period of the temperature pattern. A difficulty that is likely to arise here is that more than one periodicity arises, e.g., a large one due to a geometrical misfit and a small one due to roughness. In that case (10) or (11) ought to be applied independently to each type of irregularity. A second qualitative relationship is based on the widely accepted idea that heat in solid friction is developed only in a small fraction f of the apparent contact area. If this picture is true we get

$$[\langle(S-S_0)^2\rangle_{Av}]^{\frac{1}{2}}\sim\left(\frac{1}{f}-1\right)^{\frac{1}{2}}S_0. \qquad (13)$$

Thus we end up with the qualitative formula

$$[\langle(T-T_0)^2\rangle_{Av}]^{\frac{1}{2}}\sim\frac{1}{8}\frac{P}{k}\left(\frac{1}{f}-1\right)^{\frac{1}{2}}S_0. \qquad (14)$$

The result (14) is essentially identical with the results of a similar study by Holm.[1] Holm uses the picture of circular spots within which heat is evolved. The result to be compared with the above equation is his Eq. (13), with the explanations given by (16) and (18). In our notation we write his Eq. (16) in the form

$$K=S_0A/n.$$

Here K is the heat developed per "spot," A is the total apparent contact area and n is the total number of contact spots. In view of the model used the spot radius a is related to our f by the obvious equation

$$n\pi a^2=fA.$$

Substituting these two relations into his Eq. (13) we end up with

$$\theta_m(\infty)=\frac{1}{(\pi)^{\frac{1}{2}}}\left(\frac{A}{n}\right)^{\frac{1}{2}}\frac{S_0}{k(f)^{\frac{1}{2}}}. \qquad (15)$$

It is clear from the entire method in Holmes that he assumed that f is small compared to 1. Eqs. (14) and (15) are then essentially identical, with Eq. (15) providing a very neat concrete interpretation of our periodicity parameter P. (14) does have a smaller factor than (15) which can be partly ascribed to the fact that (14) gives a fluctuation, (15) a maximum.

Table II of the paper of Holm[1] shows that the comparison of theory and experiment is still in a tentative stage. The estimation of f by the elastic limit of the softer metal is generally accepted although unproved. More serious is the necessity of having to use for n such a small number as 1 or 3. This assumption implies essentially a geometrical misfit of the surfaces and one wonders whether the maximum temperature could not be greatly reduced by careful machining. In some observations on lubricated bearings at this laboratory profilometer readings and microscope studies seemed to point to values of P of the order 0.01 cm rather than $(A)^{\frac{1}{2}}$. This, in turn, would reduce the temperature fluctuations a hundredfold. Such a result would be significant if true.

In conclusion the author wishes to express his thanks to Dr. O. H. Clark and Dr. J. H. Buck of this laboratory for valuable discussion.

[1] Ragnar Holm, J. App. Phys. **19**, 361 (1948).

Wave Guides for Slow Waves*

L. Brillouin
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received April 26, 1948)

The paper is devoted to a discussion of the properties of electromagnetic waves propagating with a velocity $W<c$. Boundary conditions are obtained for structures with very narrow slots, either closed or open, and a general method of solution of the wave equations is discussed. Fields in the free region and in the slot region are expanded in convenient Fourier series and these series joined along the border of both regions. In doing this, it can be proved that the field contains a dominant term, completed by corrections that represent the field distortion near the open end of the slots. Under favorable conditions, and for very narrow slots, the dominant terms are by far the most important. This allows for a simplified discussion of a variety of examples. The results obtained are in agreement with those of E. L. Chu and W. W. Hansen (J. App. Phys. **18**, 996 (1947)).

I. INTRODUCTION

MANY applications of TM waves in wave guides have been recently suggested for linear accelerators of electrons. In order to obtain a good efficiency, it is necessary to match the wave velocity with electron velocity. Such a condition means that the phase-velocity of the wave must be lower than c (light velocity) while ordinary wave guides yield a phase-velocity larger than c. The present paper is devoted to a discussion of the type of guides that may satisfy this condition. Some special simplified cases are completely discussed and serve as examples of a general method that can be applied to this class of problems.

II. GENERAL FORMULAE

Let us assume that all fields depend upon t and z as

$$F(x,y,z,t)=f(x,y)\exp(i\omega t-\gamma z) \tag{1}$$

and thus correspond to a wave propagating in the z direction.

Maxwell's equations read

$$\frac{\partial E_z}{\partial y}+\gamma E_y=-i\omega\mu_0 H_x,\quad \frac{\partial H_z}{\partial y}+\gamma H_y=i\omega\epsilon_0 E_x,$$

$$-\gamma E_x-\frac{\partial E_z}{\partial x}=-i\omega\mu_0 H_y,$$

$$-\gamma H_x-\frac{\partial H_z}{\partial x}=i\omega\epsilon_0 E_y, \tag{2}$$

$$\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}=-i\omega\mu_0 H_z,\quad \frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}=i\omega\epsilon_0 E_z,$$

$$\text{div } E=0,\qquad \text{div } H=0.$$

Transverse components $E_xE_yH_xH_y$ can be expressed in terms of E_z and H_z, and Eqs. (2) yield

$$\begin{aligned}
k^2E_x&=-i\omega\mu_0\frac{\partial H_z}{\partial y}-\gamma\frac{\partial E_z}{\partial x},\\
k^2E_y&=i\omega\mu_0\frac{\partial H_z}{\partial x}-\gamma\frac{\partial E_z}{\partial y},\\
k^2H_x&=-\gamma\frac{\partial H_z}{\partial x}+i\omega\epsilon_0\frac{\partial E_z}{\partial y},\\
k^2H_y&=-\gamma\frac{\partial H_z}{\partial y}-i\omega\epsilon_0\frac{\partial E_z}{\partial x},
\end{aligned} \tag{3}$$

with

$$k^2=\omega^2\epsilon_0\mu_0+\gamma^2=\frac{\omega^2}{c^2}+\gamma^2 \tag{4}$$

Equation (3), combined with the divergence Eqs. (2) yields

$$\frac{\partial^2 E_z}{\partial x^2}+\frac{\partial^2 E_z}{\partial y^2}+k^2E_z=0, \tag{5}$$

$$\frac{\partial^2 H_z}{\partial x^2}+\frac{\partial^2 H_z}{\partial y^2}+k^2H_z=0, \tag{6}$$

* This research was supported by the Office of Naval Research and the Signal Corps, U. S. Army under ONR contract N5-ori-76, T.O. 1. A first report was issued on November 10, 1946, as Technical Report No. 4, Cruft Laboratory, Harvard University.

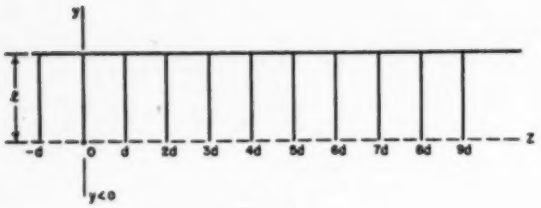


FIG. 1.

These are the fundamental equations for wave guides. Equation (5) or (6) can be solved when boundary conditions on the guide are specified. For usual guides in metallic pipes the solutions are

TE waves, $E_z=0$, H_z determined by (6)

$$\text{with} \frac{\partial H_z}{\partial n}=0 \quad (7)$$

along the boundary $\left(\frac{\partial H_z}{\partial n}\right.$ is the normal derivative$\left.\right)$.

TM waves, $H_z=0$, E_z determined by (5)

$$\text{with } E_z=0 \quad (8)$$

along the boundary.

In both cases, E_z or H_z appear as the proper functions of either (5) or (6), and correspond to some proper values $k_p^2>0$. According to (4) there appears a cut-off frequency

$$\omega_{c0}=k_p c, \quad (9)$$

for

$$\begin{array}{lll} \omega<\omega_{c0}, & \gamma \text{ is real}, & \text{attenuation}, \\ \omega>\omega_{c0}, & \gamma=i\beta=i(\omega/W), & \text{propagation}. \end{array} \quad (10)$$

In this second case from (4)

$$k^2=\omega^2\left(\frac{1}{c^2}-\frac{1}{W^2}\right)$$

or

$$\frac{1}{W^2}=\frac{1}{c^2}-\frac{k^2}{\omega^2}<\frac{1}{c^2},$$

hence,

$$W>c,$$

and the phase-velocity W is larger than c. This applies to waves in metallic pipes of any arbitrary cross section, and proves that our new condition $W<c$ cannot be satisfied by such structures.

The method followed in the next sections is to investigate the nature of waves yielding $W<c$ and to look afterward for the type of boundary that may fit the corresponding field distribution.

III. SLOW WAVES, SOME EXAMPLES

We shall assume that we have been able to obtain waves with a phase-velocity

$$W<c. \quad (11)$$

Equations (1) to (6) are still valid, but we must look for boundary conditions that may yield

$$\gamma=i(\omega/W), \quad W<c, \quad k^2=-K^2, \quad k=iK,$$

$$K^2=\omega^2\left(\frac{1}{W^2}-\frac{1}{c^2}\right)>0. \quad (12)$$

We still obtain from (3), (5), (6) the possibility to split the waves into two types, $TE(E_z=0)$ and $TM(H_z=0)$, that can be investigated separately.

TM waves correspond to the case

$$H_y=\epsilon_0 W E_x=\eta\frac{W}{c}E_x=i\omega\frac{\epsilon_0}{K^2}\frac{\partial E_z}{\partial x},$$

$$-H_x=\epsilon_0 W E_y=\eta\frac{W}{c}E_y=i\omega\frac{\epsilon_0}{K^2}\frac{\partial E_z}{\partial y}, \quad (13)$$

$$\frac{\partial^2 E_z}{\partial x^2}+\frac{\partial^2 E_z}{\partial y^2}-K^2E_z=0 \quad \eta=\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}. \quad (14)$$

Let us show on some special examples the type of solution involved. We start with plane waves

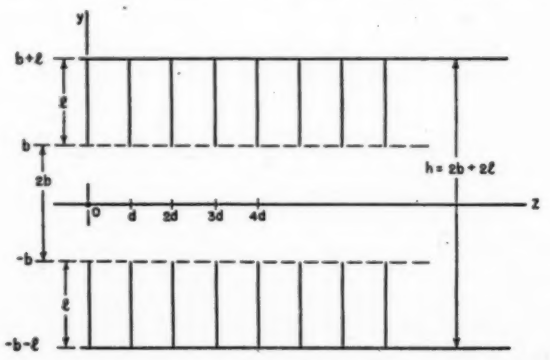
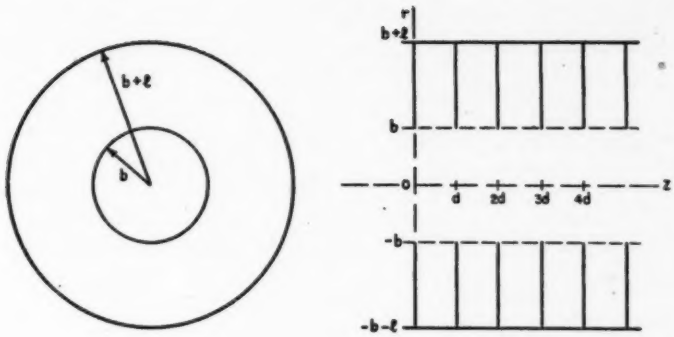


FIG. 2.

FIG. 3.



and obtain

$$E_z = e^{+Ky} \quad \text{or} \quad E_z = e^{-Ky} \tag{15a}$$

or, in a different form,

$$E_z = \cosh Ky \quad \text{or} \quad E_z = \sinh Ky. \tag{15b}$$

Solution (15a) can be used if we discover a convenient boundary that can be located on the plane $y=0$ for instance. Solution (15b) may be used between two boundaries on $y=\pm b$. Conditions on such boundaries will be discussed later (Figs. 1 and 2).

If our transmitting structure is contained between two parallel metallic planes at $x=0$ and $x=a$ we may use the following solutions:

$$E_z = \sin\frac{\pi n x}{a} e^{\pm gy}, \tag{16a}$$

$$E_z = \sin\frac{\pi n x}{a}\begin{pmatrix}\cosh\\ \sinh\end{pmatrix} gy, \tag{16b}$$

$$g^2 = K^2 + \left(\frac{\pi n}{a}\right)^2, \text{ where } n \text{ is an integer.}$$

Solutions (a) and (b) correspond to similar boundaries as for (15a, b).

For a problem with *cylindrical symmetry*, we use polar coordinates r, θ in the xy plane, and rewrite Eqs. (13) and (14) (Fig. 3):

$$H_\theta = \eta\frac{W}{c}E_r = i\omega\frac{\epsilon_0}{K^2}\frac{\partial E_z}{\partial r}, \tag{17}$$

$$-H_r = \eta\frac{W}{c}E_\theta = i\omega\frac{\epsilon_0}{K^2}\frac{1}{r}\frac{\partial E_z}{\partial \theta},$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E_z}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \theta^2} - K^2 E_z = 0, \tag{18}$$

and the solution reads

$$E_z = \cos n\theta Z_n(iKr), \tag{19}$$

where Z_n is a solution of the Bessel equation of order n. For a hollow cylindrical structure, the J_n function should be used, in order to avoid any discontinuity along the axis. For a coaxial cable or a structure with an electron beam along the axis, a combination

$$Z_n = AJ_n + BY_n \tag{20}$$

could be used in the region surrounding the beam or the central conductor.

IV. BOUNDARY CONDITIONS—DOMINANT WAVE

Starting with the simplest possible example we shall now investigate the boundary conditions for solution (15a)

$$E_z = \exp(+Ky)\,\exp(i\omega t - \gamma z) \quad y \leqslant 0, \tag{21}$$

assuming plane $y=0$ to be the boundary. Above $y=0$ we assume that we have a structure with a large number of narrow slots, with parallel metallic planes at

$$z = 0,\ d,\ 2d,\ \cdots md,\ (m+1)d \cdots \tag{22}$$

extending from $y=0$ to $y=l$, and closed at their upper end (Fig. 1). Equations (13) and (15) give the fields up to the plane $y=0$.

$$E_x = 0, \quad E_y = \frac{i\omega}{WK}\exp(Ky + i\omega t - \gamma z),$$

$$E_z = \exp(Ky + i\omega t - \gamma z), \tag{23}$$

$$H_x = -\frac{i\omega\epsilon_0}{K}\exp(Ky + i\omega t - \gamma z),$$

$$H_y = 0, \quad H_z = 0,$$

These fields will be continued, inside the slots (22), by waves propagating upwards and reflected upon the upper closed end. In order to compute these waves we rewrite Eqs. (3) to (6), taking now y instead of z, as direction of propagation:

$$
\begin{aligned}
k'^2E_z &= -i\omega\mu_0\frac{\partial H_y}{\partial x}-\gamma'\frac{\partial E_y}{\partial z},\\
k'^2E_x &= i\omega\mu_0\frac{\partial H_y}{\partial z}-\gamma'\frac{\partial E_y}{\partial x}, \quad k'^2=\frac{\omega^2}{c^2}+\gamma'^2,\\
k'^2H_z &= -\gamma'\frac{\partial H_y}{\partial z}+i\omega\epsilon_0\frac{\partial E_y}{\partial x},\\
k'^2H_x &= -\gamma'\frac{\partial H_y}{\partial x}-i\omega\epsilon_0\frac{\partial E_y}{\partial z},\\
\nabla_{xz}^2E_y+k'^2E_y&=0, \quad \nabla_{xz}^2H_y+k'^2H_y=0,
\end{aligned}
\tag{24}
$$

with a common factor $\exp(i\omega t-\gamma' y)$.

We *assume very narrow* slots, with the conditions

$$d\ll\Lambda=2\pi(W/\omega),$$

Λ wave-length in the structure, (25a)

hence

$$d\ll\lambda=2\pi(c/\omega),$$

λ wave-length in free space,

$$\omega\ll c(2\pi/d) \text{ since } W<c.$$

This condition is chosen for a first discussion because of a number of simplifications it introduces, but it is by no means necessary, and the method of solution that is presented in Sections V and VI can be generalized for any value of d and λ.

Condition (25a) also includes the following assumption

$$d\ll l \tag{25b}$$

since we shall have to deal with wave-length λ (in free space) of the order of l.

Along the open end of each slot, we may treat the fields (23) as independent of z, simply using the average z value of the slot. Let us for instance consider the slot

$$0<z<d.$$

We take the field values at $z=0$ and use formulae (23), simply assuming that the common factor $\exp(i\omega t-\gamma z)$ has been replaced by $\exp(i\omega t)$. This is a first simplification resulting from (25a).

Inside the slot we may have different types of waves propagating in the y direction, namely

$$\begin{array}{lll} TEM \text{ waves} & & E_y=H_y=0,\\ TE \text{ waves } (E_y=0), & \text{or} & TM \text{ waves } (H_y=0).\end{array}$$

Let us first discuss the TEM wave that represents the dominant term. Propagating wave would be

$$E_x=E_y=H_y=H_z=0,$$

$$E_z=\pm\frac{H_x}{\eta}=\exp\left(\mp\frac{i\omega y}{c}\right), \tag{26a}$$

but we have a reflecting wall at $y=l$ and we obtain standing waves

$$E_x=E_y=H_y=H_z=0, \tag{26b}$$

$$
\begin{aligned}
&E_z=A\sin\frac{\omega}{c}(y-l),\\
(TEM)\qquad &\\
&H_x=i\eta A\cos\frac{\omega}{c}(y-l),\\
&\eta=\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}.
\end{aligned}
$$

As for the other waves, they are given by

$$
\begin{aligned}
(TE)\quad H_y&=\sum_{m\geq 1}B_m\cos\frac{m\pi z}{d}\exp(-\gamma_m' y),\\
(TM)\quad E_y&=\sum_{m\geq 1}D_m\sin\frac{m\pi z}{d}\exp(-\gamma_m' y),
\end{aligned}
\tag{27}
$$

with

$$k_m'=\frac{m\pi}{d}, \quad \gamma_m'^2=\left(\frac{m\pi}{d}\right)^2-\frac{\omega^2}{c^2}. \tag{28}$$

The corresponding transverse fields along x and z result from Eqs. (24). These TE or TM waves (27) have the same γ_m' values, with similar cutoff frequencies

$$\omega_m=c\frac{m\pi}{d}\gg\omega. \tag{29}$$

Condition (25) shows that these cut-off frequencies are very much higher than ω, and

consequently the attenuation coefficient is very large

$$\gamma_m' \approx m\frac{\pi}{d}, \tag{28a}$$

and the waves die out, in the y direction, on a distance shorter than the width d of the slots. This represents a second simplification introduced by condition (25a).

Let us now investigate how to obtain the *continuity of the fields* along the plane $y=0$. A *TEM* wave (26b) can take care of the continuity of the E_z and H_x components that result from (23):

$$E_z = -A\sin\frac{\omega l}{c} = 1,$$
$$H_x = i\eta A\cos\frac{\omega l}{c} = -i\frac{\omega\epsilon_0}{K}, \quad y=0, \tag{30}$$

hence

$$-i\eta\frac{E_z}{H_x} = \tan\frac{\omega l}{c} = \frac{K\eta}{\omega\epsilon_0} = K\frac{c}{\omega} = \left(\frac{c^2}{W^2}-1\right)^{\frac{1}{2}}, \tag{30a}$$

a condition that determines the height l of the slot. When this is done, we are left with a single field component

$$E_y = i\frac{\omega}{WK} = \frac{i}{\left(1-\frac{W^2}{c^2}\right)^{\frac{1}{2}}}, \quad y=0, \tag{31}$$

that will excite *TM* waves (27). These waves die out on a short distance inside the slot, but they are absolutely necessary to ensure field continuity on the plane $y=0$. They will be discussed in the next section.

V. BOUNDARY CONDITIONS—SECONDARY WAVES

In order to match the residual field (31) we have to use two sets of additional fields:

I. A superposition of *TM* waves (27) inside the slots ($y>0$)

$$E_{sy} = \sum_m D_m \sin\frac{m\pi z}{d}\exp(-\gamma_m' y), \tag{32}$$

with the corresponding E_{sz} and H_{sx} field components (24), (28), (29)

$$k_m' = \frac{m\pi}{d}, \quad \gamma_m' = \left(k_m'^2 - \frac{\omega^2}{c^2}\right)^{\frac{1}{2}} \approx \frac{m\pi}{d},$$

$$E_{sz} = -\sum_m \frac{\gamma_m'}{k_m'^2}\frac{\partial E_{m,s,y}}{\partial z}$$
$$\approx -\sum_m D_m \cos\frac{m\pi z}{d}\exp(-\gamma_m' y), \tag{33}$$

$$H_{sx} = -i\omega\epsilon_0\sum_m \frac{1}{k_m'^2}\frac{\partial}{\partial z}E_{m,s,y}$$
$$= -i\omega\epsilon_0\frac{d}{\pi}\sum_m \frac{D_m}{m}\cos\frac{m\pi z}{d}\exp(-\gamma_m' y).$$

II. A superposition of waves in free space ($y<0$) according to (1), (21) and (23)

$$E_{fy} = \sum_h F_h\frac{\gamma_h}{K_h}\exp(K_h y - \gamma_h z), \tag{34}$$

$$E_{fz} = \sum_h F_h \exp(K_h y - \gamma_h z),$$
$$H_{fz} = -i\omega\epsilon_0\sum_h \frac{F_h}{K_h}\exp(K_h y - \gamma_h z),$$

with

$$\gamma_h = i\frac{\omega}{W_h}, \quad K_h^2 = \frac{\omega^2}{W_h^2} - \frac{\omega^2}{c^2} = -\gamma_h^2 - \frac{\omega^2}{c^2}.$$

The need for such a superposition of waves corresponds to a general result that has been established for any type of waves propagating through a medium possessing a periodic structure. A complete discussion of such problems may be found in the author's book.[1] Let us consider the propagation of a dominant wave, with

$$\gamma_0 = i(\omega/W) = 2\pi i a_0, \quad a_0 = 1/\Lambda,$$

where Λ is the actual wave-length in the periodic structure. Such a wave cannot propagate alone but is always accompanied by a whole set of secondary waves corresponding to

$$\gamma_h = \gamma_0 + (2\pi i h/d) \quad a_h = a_0 + (h/d) \tag{35}$$

where d is the period of the structure and h a positive or negative integer.

[1] L. Brillouin, *Wave Propagation in Periodic Structures* (McGraw-Hill Book Company, Inc., New York, 1946), pp. 5, 17 and 140.

We assumed in Section IV, Eq. (25) that the slots were very narrow

$$d \ll \Lambda.$$

This means, accordingly,

$$a_0 = 1/\Lambda \ll 1/d,$$

and

$$a_h \approx h/d, \quad \gamma_h \approx 2\pi i h/d. \tag{35a}$$

Under such conditions, the dominant wave practically yields the same field in front of a certain slot as well as in front of its nearest neighbors. This field is almost constant over a distance of a few slots. The secondary waves, on the contrary, have very short wave-lengths, d, $(\frac{1}{2})d$, $(\frac{1}{3})d \cdots (\frac{1}{h})d$, but they satisfy the condition to repeat practically the same phase on a given slot and on its nearest neighbors. This condition makes it possible to adjust the field distribution with the boundary conditions in the slots.

Comparing (35a) and (34) we obtain

$$K_h \approx (2\pi/d)|h|, \quad \gamma_h = i(h/|h|)K_h, \tag{36}$$

since condition (25a) means that $|\gamma_h|$ and K_h are much greater than ω/c. These additional free waves die out very rapidly when y increases and will contribute to the field only very near the boundary plane $y=0$. We can now write these fields explicitly.

$$\begin{aligned} E_{fy} &= i \sum_{h>0} \left\{ F_h \exp\left[\frac{2\pi h}{d}(y-iz)\right] \right. \\ &\qquad \left. - F_{-h} \exp\left[\frac{2\pi h}{d}(y+iz)\right] \right\}, \\ E_{fz} &= \sum_{h>0} \left\{ F_h \exp\left[\frac{2\pi h}{d}(y-iz)\right] \right. \\ &\qquad \left. + F_{-h} \exp\left[\frac{2\pi h}{d}(y+iz)\right] \right\}, \end{aligned} \tag{37}$$

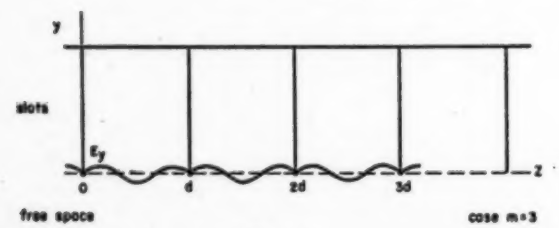


FIG. 4.

$$H_{fx} = -i\omega\epsilon_0 \frac{d}{2\pi} \sum_{h>0} \left\{ \frac{F_h}{h} \exp\left[\frac{2\pi h}{d}(y-iz)\right] + \frac{F_{-h}}{h} \exp\left[\frac{2\pi h}{d}(y+iz)\right] \right\}.$$

In order to be able to join these two different kinds of field, we have to rewrite the slot field (33) in a manner similar to expansion (37).

There would be no question for $m=2p$, but this type of slot waves is not excited since the expansion of our exciting field (31) only contains odd components:

$$E_y = i\frac{\omega}{WK} = i\frac{4}{\pi WK} \sum_p \frac{1}{2p+1} \sin(2p+1)\frac{\pi z}{d}. \tag{38}$$

So we have to discuss the case of odd m values, where we face the situation represented in Fig. 4. The expansions read

$$\begin{cases} \sin(2p+1)\pi z/d = \frac{1}{2}s_{p0} + \sum_{h\geq 1} s_{ph} \cos 2h\pi z/d \\ \qquad\qquad = \sum_{h \gtrless 0} \frac{1}{2} s_{ph} \exp\left(-2\pi i \frac{hz}{d}\right), \\ \cos(2p+1)\pi z/d = \sum_{h\geq 1} c_{ph} \sin 2h\pi z/d, \\ \qquad\qquad 0 < z < d, \end{cases} \tag{39}$$

$$s_{p-h} = s_{ph} = -\frac{4}{\pi} \frac{2p+1}{(2p+1)^2 - 4h^2},$$

$$c_{ph} = -s_{ph}\frac{2h}{2p+1} = -\frac{4}{\pi}\frac{2h}{(2p+1)^2-4h^2}.$$

Starting from expansion (32) for the slot waves and keeping only odd terms, we obtain on the plane $y=0$

$$\begin{aligned} \left\{ \begin{matrix} \text{slot waves} \\ y=0 \end{matrix} \right\} E_{sy} &= \sum_{p\geq 0} \sum_{h \gtrless 0} \tfrac{1}{2} D_{2p+1} \\ &\qquad \times s_{ph} \exp\left(-2\pi i \frac{hz}{d}\right), \\ E_{sz} &= \sum_{p\geq 0} \sum_{h \gtrless 0} \tfrac{1}{2} D_{2p+1} s_{ph} \\ &\qquad \times \frac{2h}{2p+1} \exp\left(-2\pi i \frac{hz}{d}\right), \\ H_{sx} &= \frac{\omega\epsilon_0 d}{2\pi} \sum_{p\geq 0} \sum_{h \gtrless 0} \\ &\qquad \times \frac{D_{2p+1} 2h}{(2p+1)^2} \exp\left(-2\pi i \frac{hz}{d}\right). \end{aligned} \tag{40}$$

Along the same plane $y=0$, the free waves yield[2]

$$\left\{\begin{matrix}\text{free waves}\\ y=0\end{matrix}\right\} \quad F_0=0,$$

$$E_{fy}=i\sum_{h\gtrless 0}F_h\frac{h}{|h|}\times\exp\left(-2\pi i\frac{hz}{a}\right), \quad (41)$$

$$E_{fz}=\sum_{h\gtrless 0}F_h\exp\left(-2\pi i\frac{hz}{d}\right),$$

$$H_{fz}=-i\frac{\omega\epsilon_0 d}{2\pi}\sum_{h\gtrless 0}\frac{F_h}{|h|}\times\exp\left(-2\pi i\frac{hz}{d}\right).$$

The total field resulting from slot and free waves must be equal to our residual field (31); hence

$$E_{sy}+E_{fy}=i\frac{\omega}{WK}, \quad E_{sz}+E_{fz}=0, \quad H_{sz}+H_{fz}=0. \quad (42)$$

We may ignore the last condition since our magnetic fields H_{sz} and H_{fz} are of the order of magnitude d and can be neglected, according to conditions (25) or (29), and we are left with the set of equations

$$h\neq 0, \quad i\frac{h}{|h|}F_h+\tfrac{1}{2}\sum_{p\geq 0}D_{2p+1}s_{ph}=0,$$

$$F_h+\frac{i}{2}\sum_{p\geq 0}D_{2p+1}s_{ph}\frac{2h}{2p+1}=0,$$

$$h=0, \quad \tfrac{1}{2}\sum_{p\geq 0}D_{2p+1}s_{p0}=i(\omega/WK).$$

Let us consider separately the cases of positive and negative h:

$$h>0, \quad F_h=-\frac{i}{2}\sum_{p\geq 0}D_{2p+1}s_{ph}\frac{2h}{2p+1} \quad (44)$$

$$=\frac{i}{2}\sum_{p\geq 0}D_{2p+1}s_{ph}$$

$$=-F_{-h},$$

since $s_{ph}=s_{p-h}$. Both expressions (44) must be equal, and we are left with the following equations for the D coefficients.

FIG. 5.

$$h>0, \quad \sum_{p\geq 0}D_{2p+1}\frac{2p+1+2h}{2p+1}s_{ph}=\sum_{p\geq 0}D_{2p+1}\frac{4}{\pi}\frac{1}{2p+1-2h}=0, \quad (45)$$

$$h=0, \quad \tfrac{1}{2}\sum_{p\geq 0}D_{2p+1}s_{ph}=\frac{2}{\pi}\sum_{p\geq 0}\frac{D_{2p+1}}{2p+1}=i\frac{\omega}{WK}.$$

When the D_{2p+1} coefficients are obtained, the corresponding F_h values result from (44).

VI. MATHEMATICAL SOLUTION OF EQUATIONS

Let us introduce an analytic function of a continuous variable θ with the following definition:

$$f(\theta)=\sum_{p\geq 0}\frac{d_{2p+1}}{2p+1-2\theta}, \quad (46)$$

and require that

$$f(0)=1, \quad f(h)=0, \quad h>0 \text{ integer}. \quad (47)$$

The function $f(\theta)$ has *poles* at

$$\theta=p+\tfrac{1}{2}, \quad p\geq 0, \quad (48)$$

and the corresponding residues are $-\frac{1}{2}d_{2p+1}$. If we can build the function f, then compute its residues d_{2p+1}, the solution of (45) obtains

$$D_{2p+1}=i\frac{\pi}{2}\frac{\omega}{WK}d_{2p+1}. \quad (49)$$

Looking at conditions (47), (48), we immediately recognize their similarity with the properties of

$$\tan\pi\theta/\pi\theta, \quad (50)$$

but this last function has poles for positive or negative p values, while we want only positive values to come in. In order to cancel all the poles

[2] The condition $F_0=0$ means that we do not want to reintroduce any component of the dominant wave. It was implicitly stated in Eq. (37).

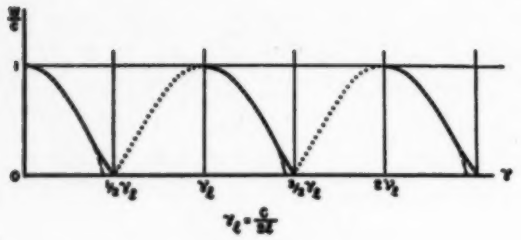


FIG. 6.

on the negative side, we may use a function $\Gamma(\frac{1}{2}+\theta)$ that has poles at

$$\theta=-\frac{1}{2},\quad -\frac{3}{2},\quad -\frac{5}{2},$$

and take

$$f(\theta)=\frac{(\pi)^{\frac{1}{2}}}{\Gamma(\frac{1}{2}+\theta)}\frac{\tan\pi\theta}{\pi\theta}. \tag{51}$$

This function satisfies all our conditions (46), (47), (48), and equals 1 when $\theta=0$, since $\Gamma(\frac{1}{2})=(\pi)^{\frac{1}{2}}$. Thus, it represents the only possible analytic function of this type. Its general behavior is shown in Fig. 5. The residues are easily computed since for

$$\theta=p+\tfrac{1}{2}+\epsilon,\quad \tan\pi\theta\approx\frac{-1}{\pi-\epsilon},$$

$$\Gamma(p+1)=p!,$$

and

$$d_{2p+1}=\frac{2}{\pi^{\frac{1}{2}}(p+\frac{1}{2})p!},$$

$$d_1=\frac{4}{\pi^{\frac{1}{2}}}\approx 0.717, \tag{52}$$

$$d_3=\tfrac{1}{3}d_1,$$

$$d_5=\tfrac{1}{10}d_1.$$

Hence, according to Eq. (49)

$$D_{2p+1}=\frac{i}{(\pi)^{\frac{1}{2}}(p+\frac{1}{2})p!}\frac{\omega}{WK}, \tag{53}$$

and that solves our problem. The F_h coefficients result from (44), (39):

$$F_h=\frac{i}{2}\sum_{p\geq 0}D_{2p+1}\frac{4}{\pi}\frac{2p+1}{(2p+1)^2-4h^2}$$

$$=-\sum_{p\geq 0}\frac{4}{\pi^{\frac{1}{2}}p![(2p+1)^2-4h^2]}\frac{\omega}{WK}. \tag{54}$$

These sums can be computed numerically and amount to a small fraction of ω/WK.

VII. PHYSICAL PROPERTIES OF THE WAVES

The dominant wave defined by Eqs. (12) and (30) propagates with a velocity W that was originally assumed to be smaller than c, but under certain conditions it may also occasionally become larger than c, as will be seen later.

In our present problem, we obtain (Eq. (30a))

$$\tan\frac{\omega l}{c}=K\frac{c}{\omega}=\left(\frac{c^2}{W^2}-1\right)^{\frac{1}{2}}. \tag{55}$$

This defines the dominant wave, with a phase velocity

$$W/c=1\Big/\left(1+\tan^2\frac{\omega l}{c}\right)^{\frac{1}{2}}=|\cos\omega l/c|. \tag{56}$$

But there is a limitation to the possible use of these formulae. Our theory *works only when* $K>0$, that is, when free waves die out exponentially away from the plane boundary $y=0$. Negative K values would yield waves increasing exponentially when $y\to-\infty$, and corresponding to no possible physical solution.

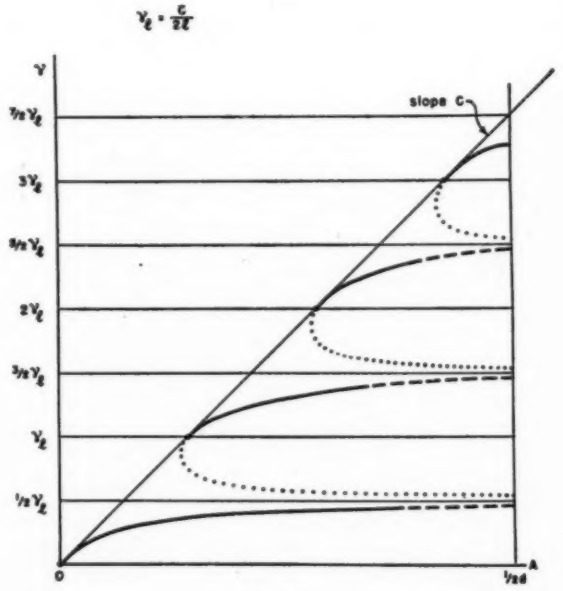


FIG. 7.

This very important remark shows that we obtain *no solution* when

$$\frac{\pi}{2}<\frac{\omega l}{c}<\pi, \qquad \frac{3\pi}{2}<\frac{\omega l}{c}<2\pi,$$

$$\frac{1}{2}\nu_l<\nu<\nu_l, \qquad \frac{3}{2}\nu_l<\nu<2\nu_l, \qquad \nu_l=\frac{c}{2l}.$$

This is shown on Figs. 6 and 7 where the meaningless parts of the curves have been drawn as dotted curves. Figure 6 shows the variation of the phase velocity W as a function of frequency ν. Figure 7 is a graph of ν as a function of the inverse wave-length A.

$$A=1/\Lambda=\nu/W. \tag{58}$$

The limits $1/2\nu_l$, $3/2\nu_l$ correspond to resonance in the slots ($\lambda=4l$, $4/3l$) while the other limits ν_l, $2\nu_l$, mean anti-resonance and $\lambda=2l, l, \cdots$ The theory developed in the preceding sections is based upon the assumption (25a)

$$\omega/W=2\pi A\ll 2\pi/d,$$

but it is easy to foresee what happens when A becomes larger and approaches $1/2d$. The curve $\nu(A)$ will reach the limit with an horizontal tangent, as shown on Fig. 7. Large values of A are obtained whenever the phase-velocity becomes very small; hence the curve of Fig. 7 will be slightly distorted in its lower parts (Fig. 6 dashed curve). The group velocity U is given by

$$U=\left|\frac{d\nu}{dA}\right|, \tag{59}$$

as shown in the author's book[3] and has been

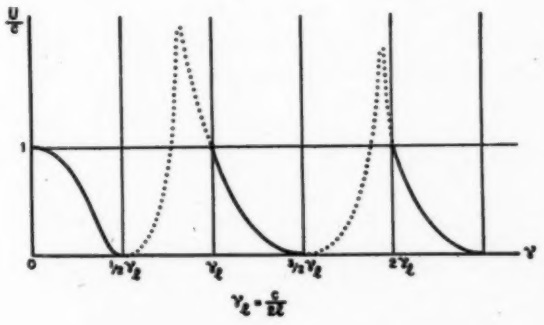


FIG. 8.

[3] See reference 1, Chap. I, II, IX, and V.

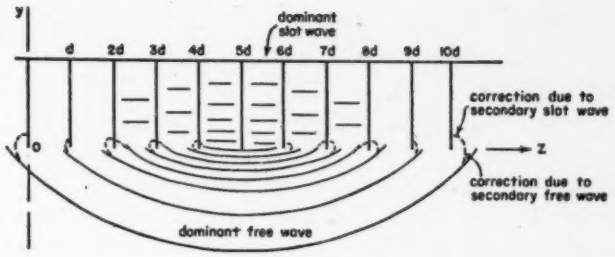


FIG. 9.

plotted in Fig. 8. We have

$$A=\nu\Big/\left(c\left|\cos\frac{\omega l}{c}\right|\right)$$

$$\frac{dA}{d\nu}=\frac{\left|\cos\frac{\omega l}{c}+\frac{\omega l}{c}\sin\frac{\omega l}{c}\right|}{c\left(\cos\frac{\omega l}{c}\right)^2}. \tag{60}$$

This expression becomes zero when $\omega l/c\,\tan\omega l/c=-1$, and this would mean U infinite, a situation that occurs only along the forbidden branches of curve ν.

The role played by the secondary waves is explained on Fig. 9. The "free" waves extend a short distance below the "boundary" $y=0$ and contribute to bending the lines of forces just below this boundary. The secondary slot waves achieve the correction and bend the lines of forces until they reach the metal plates at right angles. Both free and slot waves exhibit a considerable amount of attenuation when $|y|$ increases. The most important (and more far-reaching) correction in free space is given by the first secondary wave, whose amplitude is approximately 1/10 of that of the dominant free wave, and whose attenuation is in $(\exp(2\pi/d)y)$ (Eq. (36)).

This description is correct so long as the dominant wave has a large wave-length $\Lambda\gg d$, but we notice that serious modification ought to be introduced in the theory when Λ approaches $2d$ (Eq. (58), Figs. 6 and 7). In this case, the so-called dominant free wave and the first secondary free wave become of comparable intensity and attenuation, and our general approximations should be revised.

As a rule there is a considerable accumulation of lines of forces on the sharp edges of the fins that separate successive slots. This very large field should be logarithmically infinite, and would show up in the amplitudes of all upper terms, as well in slot waves as in free waves.

The complete theory developed in the preceding sections should be very useful in showing the importance of the field perturbation near the slots, and computing the distance at which the dominant free wave can be considered as practically unperturbed.

VIII. WAVES BETWEEN TWO PARALLEL PLANE SLOT SYSTEMS

Symmetrical *TM* Waves

The case of one single plane boundary is of great theoretical interest, and its discussion will enable us to investigate now the problem of waves propagating between two parallel plane boundaries, a structure of greater practical significance for possible experimentation. The structure is shown in Fig. 2 and its boundaries are at $y=\pm b$, while the slots have a height l. We may obtain two different types of waves in free space ($-b\leqslant y\leqslant b$) according to the solutions (15b) used for the dominant TM wave:

symmetrical wave

$$E_z=\cosh Ky\,\exp(i\omega t-\gamma z), \quad (61a)$$

antisymmetrical

$$E_z=\sinh Ky\,\exp(i\omega t-\gamma z). \quad (61b)$$

The case of a dominant TEM wave should also be discussed. Let us discuss first the *symmetrical TM wave* that is the most important one for application since it yields a longitudinal field along the medium plane $y=0$. Its complete field distribution is obtained from Eq. (13):

$$E_z=\cosh Ky\,\exp(i\omega t-\gamma z),$$

$$E_y=\frac{\gamma}{K}\sinh Ky\,\exp(i\omega t-\gamma z),$$

$$H_x=-\frac{i\omega\epsilon_0}{K}\sinh Ky\,\exp(i\omega t-\gamma z).$$

We assume the same sort of dominant slot wave as in Section IV (Eq. (26b)) and state that this first slot wave must insure continuity of the E_z and H_x fields along the plane $y=b$. This leads to a condition similar to (30):

$$\frac{H_x}{E_z}=-\frac{i\omega\epsilon_0}{K}\tanh Kb=\frac{-i\eta}{\tan(\omega l/c)}=\frac{-i\omega\epsilon_0}{k'\tan k'l}, \quad (62)$$

with $k'=\omega/c$ at $y=b$, and a similar condition for $y=-b$. Condition (62) leads to the following relation:

$$B\coth B=(b/l)\Omega\tan\Omega,$$

where $B=Kb$,

$$\Omega=k'l=\omega l/c=2\pi(l/\lambda), \quad (63)$$

and we have to distinguish between cases where B is real and cases where $B=i\beta$, when $B\coth B$, is replaced by $\beta\cot\beta$.

Case I, B real, $(b/l)\Omega\tan\Omega=B\coth B\geq 1$,

$$K=B/b \text{ real}, \qquad K^2=\omega^2\left(\frac{1}{W^2}-\frac{1}{c^2}\right), \quad (64)$$

$$\frac{c^2}{W^2}=1+K^2\frac{c^2}{\omega^2}>1, \quad W/c<1.$$

The lowest value 1 is obtained for $B\ll 1$.

Case II, $B=i\beta$, $(b/l)\Omega\tan\Omega=\beta\cot\beta$

$$K=i\frac{\beta}{b}=ik, \qquad k^2=\omega^2\left(\frac{1}{c^2}-\frac{1}{W^2}\right), \quad (65)$$

$$\frac{c^2}{W^2}=1-k^2\frac{c^2}{\omega^2}<1, \quad \frac{W}{c}>1.$$

Since $\beta\cot\beta$ can take any value from $-\infty$ to $+\infty$, we need an additional condition to tell us when to use case I or case II. This results from the fact that our expression for c^2/W^2 must be positive; hence the only acceptable values for k are those for which

$$0\leqslant k\frac{c}{\omega}\leqslant 1, \quad 0\leqslant\frac{kc}{\omega}=\frac{k}{k'}=\frac{\beta l}{\Omega b}\leqslant 1. \quad (66)$$

The situation is better discussed graphically by using the graphs of Fig. 10. Starting from low frequencies, small Ω, we first do not obtain any solution since β should be nearly $\pi/2$ and β/Ω very large, hence (66) cannot hold, or to put it more clearly, the solution needs complex B

and K values and yields an attenuated wave instead of propagation. Low frequencies are in a stopping band. The lower limit of the passing band is obtained when condition (66) is first satisfied:

First cut-off Ω_1,

$$\frac{\beta_1 l}{\Omega_1 b}=1, \quad \cot\beta_1=\tan\Omega_1, \tag{67}$$

hence

$$\beta_1=\frac{\pi}{2}-\Omega_1,$$

and

$$\frac{\beta_1}{\Omega_1}=\frac{\pi}{2\Omega_1}-1=\frac{b}{l}, \quad \Omega_1=\frac{\pi}{2}\frac{l}{b+l}, \quad \text{Case II.}$$

This corresponds, according to (63), to a wavelength λ_1:

$$\lambda_1=\frac{2\pi c}{\omega_1}=\frac{2\pi l}{\Omega_1}=4(b+l)=2h, \quad h=2b+2l. \tag{68}$$

Starting from these points $\beta_1\Omega_1$ marked by circles on the graphs of Fig. 10, and increasing the frequency, one moves up along both curves, until another limit is reached when

$$\beta_2=0, \quad \beta_2\cot\beta_2=1, \quad \Omega_2\tan\Omega_2=\frac{l}{b}. \tag{69}$$

The points are marked by crosses on both curves. When Ω now increases above Ω_2 we must switch to Case I, Eq. (64) and take B real, and this will work until we reach another limit at

$$\begin{aligned} &B\to\infty, \quad \Omega_3\tan\Omega_3\to\infty, \quad \Omega_3=\pi/2, \\ &\omega_3=(\pi/2)(c/l), \quad \lambda_3=4l. \end{aligned} \tag{70}$$

This is the upper cut-off for the first passing band.[4]

The second passing band starts as soon as condition (66) is satisfied again, with a situation similar to (67).

$$\begin{aligned} &\frac{\beta_4 l}{\Omega_4 b}=1, \qquad \cot\beta_4=\tan\Omega_4, \quad \beta_4=\frac{3\pi}{2}-\Omega_4, \\ &\Omega_4=\frac{3\pi}{2}\frac{l}{b+l}, \quad \lambda_4=\frac{2}{3}h, \qquad \text{Case II.} \end{aligned} \tag{71}$$

[4] Comparing (68) and (70) it will be noted that this discussion corresponds to the case $\omega_1<\omega_3$; hence $l<b$. The band structure would be different in the case $b<l$.

This is the lower cut-off of the second passing band. We have to switch to Case I when Ω reaches Ω_3 which brings β back to zero again, and the upper cut-off is reached when

$$\Omega_6=\frac{3\pi}{2}, \quad \omega_6=\frac{3\pi c}{2l}, \quad \lambda_6=\frac{4l}{3}, \tag{72}$$

and the same process is repeated again in the upper branches.

The preceding discussion is based upon the consideration of dominant waves both in free space and in the slot. But in addition to these waves, we will need an infinite set of free and slot waves to complete the matching of fields along the boundary $y=\pm b$. This will be done along the same line as the discussion of Sections V and VI, but special attention should be paid to conditions of Case II, when the secondary waves in free space may extend throughout the free interval $\pm b$. In Case I, secondary free waves will die exponentially and be of importance only near the slot where they are needed to correct the shape of the lines of forces, as shown in Fig. 11. The problem of these secondary waves

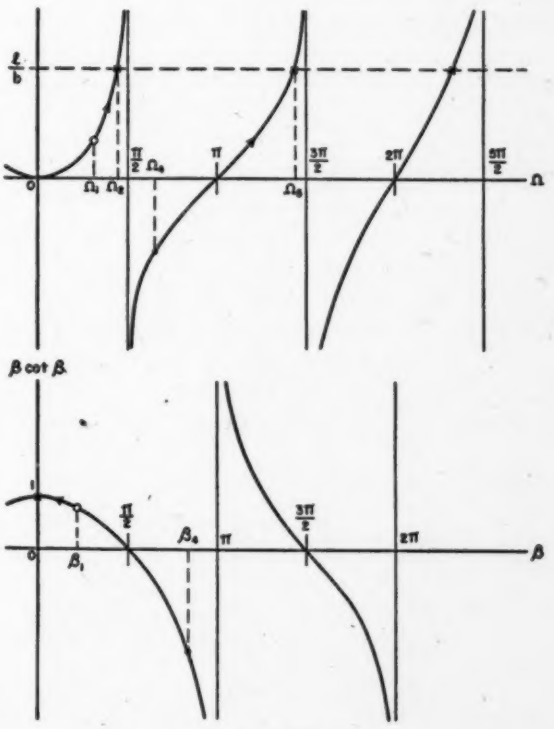


FIG. 10.

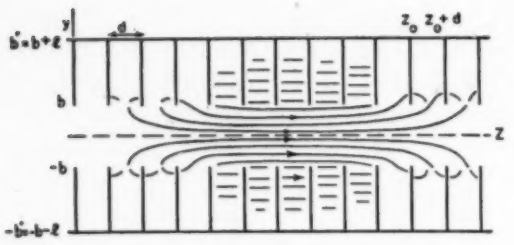


FIG. 11.

and of their possible reaction on the dominant waves would require special discussion.

Our discussion was based upon the assumption

$$\nu_h < \tfrac{1}{2}\nu_l, \quad h = 2b + 2l > 4l, \quad l < b, \qquad (73)$$

and corresponding curves were plotted in Figs. 12 and 13. If condition (73) were not realized, the shape of the curves would be different. Figure 13 shows the wave velocity as a function of frequency. The group velocity is always smaller than c. Only the first two branches have been plotted but there would be an infinite set of similar passing bands for other frequencies.

The results here obtained bear a striking similarity to those of Kelvin and Vincent for a mechanical structure exhibiting just one resonance frequency[5] and yielding two passing bands, with a stopping band in between, near the resonance frequency. Here we have an infinite series of resonance frequencies in the slot, with $\nu = 1/2\nu_l,\ 3/2\nu_l,\ 5/2\nu_l \cdots$ and we obtain an infinite set of alternating passing and stopping bands.

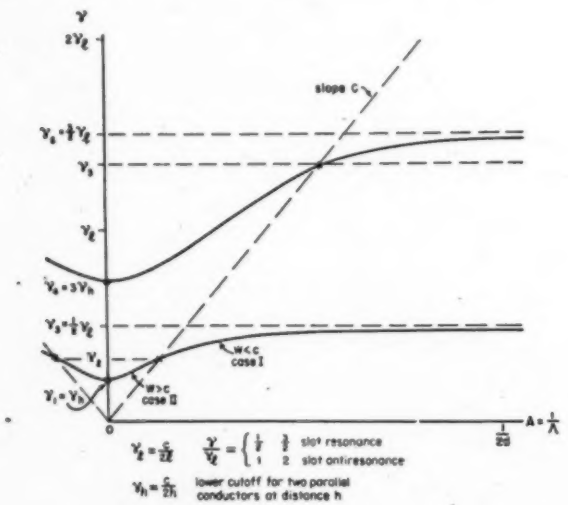


FIG. 12.

[5] See reference 1, pp. 11 and 12, Figs. 3.2, 3.3 and 3.4.

Antisymmetrical *TM* Waves

Symmetrical waves discussed in the preceding section are the most important ones for practical application to electron acceleration or deceleration, since they yield a strong longitudinal electric field on the middle plane $y=0$. In addition to these waves, we found another set of TM antisymmetrical waves (61b) that we propose now to discuss. Their field distribution is given by

$$E_z = \sinh Ky \exp(i\omega t - \gamma z),$$

$$E_y = \frac{\gamma}{K} \cosh Ky \exp(i\omega t - \gamma z), \qquad (74)$$

$$H_x = -\frac{i\omega\epsilon_0}{K} \cosh Ky \exp(i\omega t - \gamma z).$$

We follow the same procedure as in Section VIII and obtain, instead of (62), the following condition for the junction of the dominant waves in free space and in the slot along the plane $y=b$.

$$\frac{H_x}{E_z} = -\frac{i\omega\epsilon_0}{K \tanh Kb} = -\frac{i\omega\epsilon_0}{k' \tan k'l}, \quad k' = \frac{\omega}{c}, \qquad (75)$$

and relations (63) and (64) are now replaced by

$$\text{I} \quad B = Kb > 0 \text{ real}, \quad \Omega = k'l = \omega\frac{l}{c}, \qquad (76a)$$

$$-\frac{b}{l}\Omega \tan\Omega = B \tanh B \geqslant 0,$$

$$\text{II} \quad B = Kb = i\beta,$$

$$-\frac{b}{l}\Omega \tan\Omega = -\beta \tan\beta, \qquad (76b)$$

with an additional condition similar to (66):

$$0 \leqslant \beta l/\Omega b \leqslant 1. \qquad (77)$$

The discussion is slightly different from the one given in Section VIII (Eqs. (67) to (72)) for symmetrical waves. Starting with low frequencies, Ω small, we may immediately use Case I (Eq.

(76a)) and we obtain

$$\Omega \ll 1, \quad \tan\Omega \approx \Omega, \quad \tanh B \approx B, \quad (b/l)\Omega^2 = B^2. \quad (78)$$

For these low frequencies the wave is almost *TEM* since K is very small and $E_z = \sinh Kb$ is practically naught. The velocity of these waves is smaller than c but still rather high since (Eq. (64)) yields

$$\Omega \ll 1, \quad \frac{c^2}{W^2} = 1 + K^2\frac{c^2}{\omega^2} = 1 + \left(\frac{Bl}{b\Omega}\right)^2 \approx 1 + \frac{l}{b}. \quad (79)$$

When Ω increases, B does the same and the limit is found when

$$\Omega = \pi/2, \quad B = \infty, \quad \omega = \omega_3 = \frac{\pi c}{2l}, \quad \nu_3 = \frac{1}{2}\nu_l, \quad (80)$$

that is identical with the upper cut-off ω_3 of the first passing band for symmetrical waves (Section VIII), and corresponds to resonance inside the slots. When Ω gets larger than $\pi/2$, $\Omega\tan\Omega$ becomes negative and we must switch to Case II, Eq. (76b) using a value just *below* $\pi/2$, so as to make $\beta\tan\beta > 0$.

According to condition (77) such a solution can always be built if $l < b$ (starting from $\Omega = \pi/2 + \epsilon_1$, $\beta = \pi/2 - \epsilon_2$) but there is a limitation if $l > b$ and the lower cut-off of the second passing band appears when

$$\Omega b = \beta l, \quad (\text{case } l > b) \quad (81)$$

and condition (77) is just fulfilled. This means

$$\tan\Omega = -\tan\beta, \quad \Omega = \pi - \beta, \quad \beta < \frac{\pi}{2}, \quad \Omega > \frac{\pi}{2}.$$

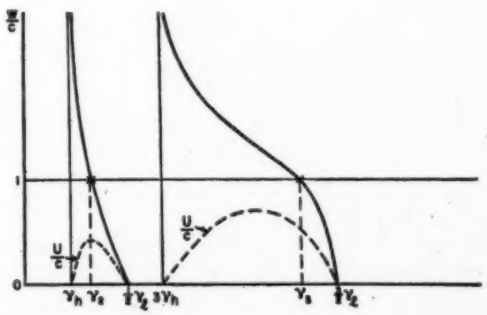


FIG. 13.

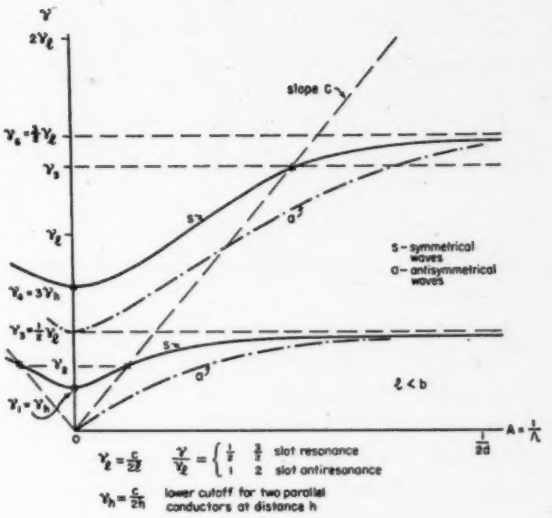


FIG. 14.

Hence

$$\frac{B}{\Omega} = \frac{\pi}{\Omega} - 1 = \frac{b}{l}, \quad \Omega = \pi\frac{l}{b+l},$$

$$\lambda = \frac{2\pi l}{\Omega} = 2(b+l) = 2h,$$

$$\nu = \nu_h > \nu_3 = \tfrac{1}{2}\nu_l. \quad (82)$$

As in Section VIII, we choose the case $l < b$ and all our diagrams have been plotted according to this assumption.

Increasing Ω from $\pi/2$ to $3\pi/2$ we note that for $\Omega = \pi$ we obtain $\beta = 0$, and for $\pi < \Omega < 3\pi/2$ we must switch back to Case I. The upper limit is

$$\Omega = 3\pi/2, \quad B = \infty, \quad \omega = \omega_b, \quad (83)$$

and coincides with the upper limit of the second passing band for symmetrical waves (Section

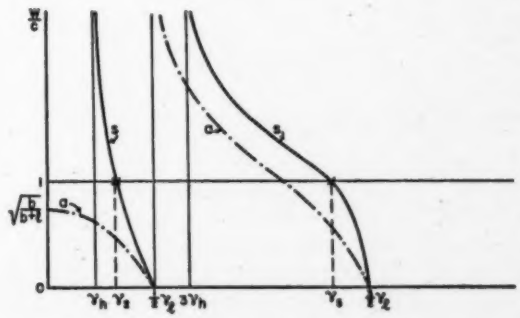


FIG. 15.

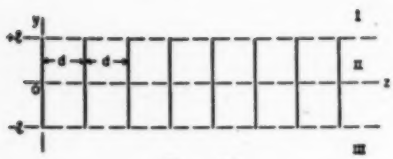


FIG. 16.

VIII). The reason is the same as in the case of Ω_3 (Eq. (80)), since both cut-offs are due to the second resonance in the slots.

The corresponding curves have been plotted in Figs. 14 and 15 that correspond to Figs. 12 and 13 and where both types of solutions have been plotted together, the letter "*a*" corresponds to the new antisymmetrical solution, while "*s*" indicates the former symmetrical solution of Figs. 12 and 13.

IX. PLANE STRUCTURE WITH OPEN SLOTS

The method inaugurated in the first sections (I to VI) can be used for some other problems and we shall now investigate the case of a plane structure with open slots, such as the one shown in Fig. 16. Wave propagation about such a structure will contain three different sets of waves:

$$\begin{array}{ll} \text{waves in free region I} & l \leqslant y, \\ \text{waves inside slots II} & -l \leqslant y \leqslant l, \\ \text{waves in free region III} & y \leqslant -l. \end{array} \tag{84}$$

Our first type of solution will consist of *antisymmetrical waves*, with opposite fields on both sides and no E_z field along the medium plane $y=0$. This will be exactly twice the solution for a closed slot, since a conducting plane could be introduced along the plane $y=0$ without causing any disturbance. Hence this first type of wave is exactly the same as the one investigated in Sections IV, V, VI and VII. Inside the slots, the solution is in

$$E_z = A \sin k'y, \quad k' = \omega/c,$$

and more generally

$$E_z(y) = -E_z(-y), \tag{85}$$

so that the E_z field in region I above the slot system, for instance, is just opposite to the E_z field in region III below the slot. The properties of these antisymmetrical waves are summarized Figs. 6, 7 and 8.

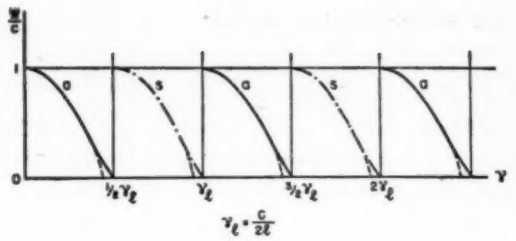


FIG. 17.

In addition to these solutions, we obtain another set of *waves* with *symmetrical* structure, when

$$E_z(y) = E_z(-y), \tag{86}$$

resulting in the *slots* in a field distribution

$$\begin{aligned} E_z &= A \cos k'y, \\ E_y &= 0, \\ H_x &= -i\omega\epsilon_0 \frac{A}{k'} \sin k'y, \end{aligned} \tag{87}$$

for the dominant wave (*TEM*), instead of (26b). According to condition (86) we need study the field in only one of the two regions I and III. Let us take region III and a dominant wave similar to that already used in Section IV (Eqs. (21) and (23)). The next step is to write a condition similar to (30), insuring continuity of E_z and H_x along the plane $y=-l$, and ignoring the E_y

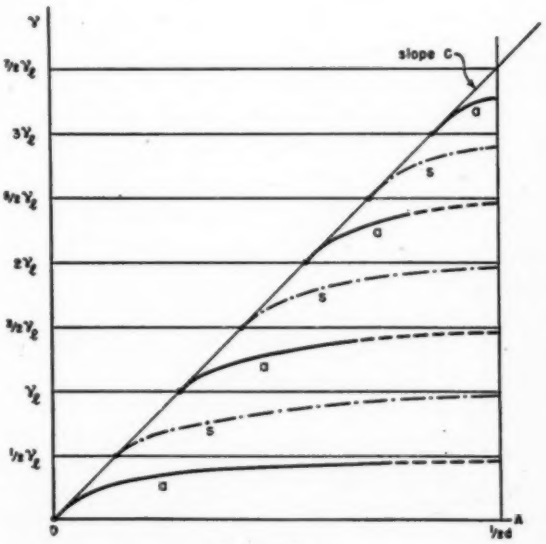


FIG. 18.

field that would subsequently be taken care of along the lines of Sections V and VI. The continuity of E_z and H_z is obtained when the ratio

$$\frac{H_z}{E_z}=-\frac{i\omega\epsilon_0}{K}=+\frac{i\omega\epsilon_0}{k'}\tan k'l \quad (y=-l) \tag{88}$$

has the same value in both regions; hence the preceding equation, in which the first fraction corresponds to waves in region III (Eq. (23)) and the second fraction to the slot waves (87). For symmetrical waves, Eq. (88) plays the same role as Eq. (30a) for antisymmetrical waves, and yields

$$K=-k'\cot k'l>0, \quad k'=\frac{\omega}{c}, \tag{89}$$

since a positive K is needed to yield waves decreasing exponentially at large distance $|y|$.

According to (12) we have again

$$\frac{K^2}{k'^2}=K^2\frac{c^2}{\omega^2}=\frac{c^2}{W^2}-1.$$

Hence in our problem

$$\frac{c^2}{W^2}=1+\frac{K^2}{k'^2}=1+\cot^2 k'l=\frac{1}{\sin^2 k'l},$$

$$\frac{W}{c}=\left|\sin\frac{\omega l}{c}\right|, \tag{90}$$

(antisymmetrical waves) instead of formula (56) with a $|\cos\omega l/c|$ for the symmetrical waves. Condition (89), that K must be positive, yields the *passing bands*

$$\frac{\pi}{2}\leqq\frac{\omega l}{c}\leqq\pi, \quad \frac{1}{2}\nu_l\leqq\nu\leqq\nu_l,$$

$$\frac{3\pi}{2}\leqq\frac{\omega l}{c}\leqq 2\pi, \quad \frac{3}{2}\nu_l\leqq\nu\leqq 2\nu_l,$$

and the diagrams of Figs. 7, 8 and 9 are replaced by those of Figs. 17, 18 and 19 where both types of waves have been plotted together, the letter "a" meaning our new antisymmetrical solution, and "s" corresponding to the former symmetrical solution. Here again our solution assuming a long wave-length compared to the slot thickness d should be corrected as indicated by the dashed lines.

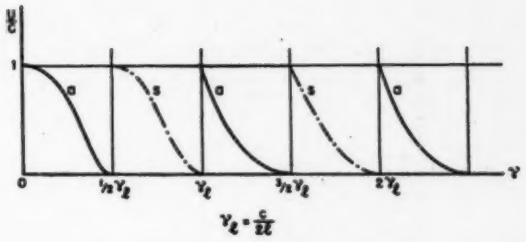


FIG. 19.

X. TWO PARALLEL OPEN-SLOT PLANE SYSTEMS

The structure is represented in Fig. 20 and is similar to the one investigated in Sections VIII and IX (Fig. 2) but for the fact that we now have open slots instead of closed ones. We are again interested in TM waves and we have to distinguish between

$$\begin{array}{ll}\text{symmetrical waves} & E_z(y)=E_z(-y),\\ \text{antisymmetrical waves} & E_z(y)=-E_z(-y).\end{array} \tag{92}$$

The waves are represented by different functions in the successive regions

I	Above the structure	$b+l<y$,
II	In the upper slot	$b<y<b+l$,
III	Free region between slots	$-b<y<b$.

Regions IV and V in and below the lower slots do not need special discussion since the corresponding waves result from the solutions obtained in II and I, depending upon the type of symmetry (92) of the wave.

Let us start with *symmetrical waves*. In region III we use again a wave of type (61c) as in Section VIII. In the slots we use standing waves of

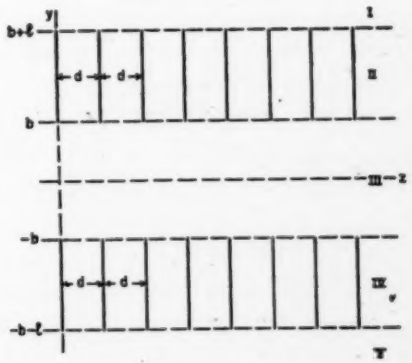


FIG. 20.

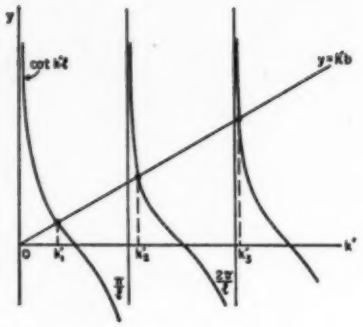


FIG. 21.

the general type (26b), but with an unknown phase φ:

$$\begin{aligned} E_z &= A\cos(k'y+\varphi), \\ E_y &= 0, \\ H_x &= -i\eta_0 A\sin(k'y+\varphi), \\ k' &= \frac{\omega}{c}, \quad \eta_0 = \frac{\omega\epsilon_0}{k'} = \epsilon_0 c. \end{aligned} \tag{93}$$

Finally we come to region I above the slots and use an exponential wave

$$\begin{aligned} E_z &= Be^{-Ky}, \\ E_y &= -\frac{i\omega}{WK}Be^{-Ky} \\ H_x &= \frac{i\omega\epsilon_0}{K}Be^{-Ky}, \end{aligned} \tag{94}$$

similar to (13), (15) and (23) but with a $-K$ instead of $+K$ since we want the wave to decrease to zero when y increases to infinity. All waves have a common factor in $\exp(i\omega t-\gamma z)$. K is the same in (94) and (61c) since γ is the same for all waves, (Eq. (12)). Now we have to write the continuity of E_z and H_x along both planes $y=b$ and $y=b+l$,

$$y=b, \quad \frac{H_x}{E_z} = -i\omega\epsilon_0\frac{\tanh Kb}{K} = -i\omega\epsilon_0\frac{\tan(k'b+\varphi)}{k'}, \tag{95}$$

by comparison of (61c) and (93)

$$y=b+l, \quad \frac{H_x}{E_z} = \frac{i\omega\epsilon_0}{K} = -\frac{i\omega\epsilon_0}{k'}\tan(k'b+k'l+\varphi), \tag{96}$$

by comparison of (93) and (94). Let us put

$$k'b+\varphi=\psi,$$

and we have

$$\frac{\tanh Kb}{K} = \frac{\tan\psi}{k'}, \quad -\frac{1}{K} = \frac{\tan(\psi+k'l)}{k'}. \tag{97}$$

Calling

$$T=\tanh(Kb), \quad t=\tan(k'l),$$

we obtain

$$\tan(\psi+k'l) = \frac{\tan\psi+t}{1-t\cdot\tan\psi} = \frac{(k'/K)T+t}{1-(k/'K)tT} = -\frac{k'}{K},$$

or

$$-T\frac{k'}{K} = \left(t+\frac{k'}{K}\right)\Big/\left(1-\frac{k'}{K}t\right). \tag{98}$$

Since K must be positive, T varies from 0 (when $K=0$) to 1, when K becomes infinite. We obtain the solutions corresponding to the limiting cases by taking first

K very small, $T=Kb$,
$-T(k'/K) = -k'b$, k'/K very large,

and Eq. (98) reduces to

$$-k'b = -\frac{1}{t} = -\cot k'l. \tag{99}$$

That gives the values k_1', k_2', k_3' as shown in Fig. 21. Next we take

$$K \text{ very large}, \quad T=1, \quad \frac{k'}{K}\approx 0,$$

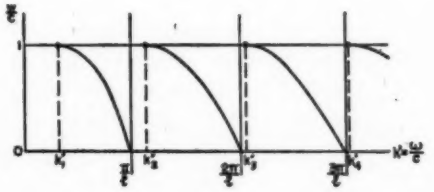


FIG. 22.

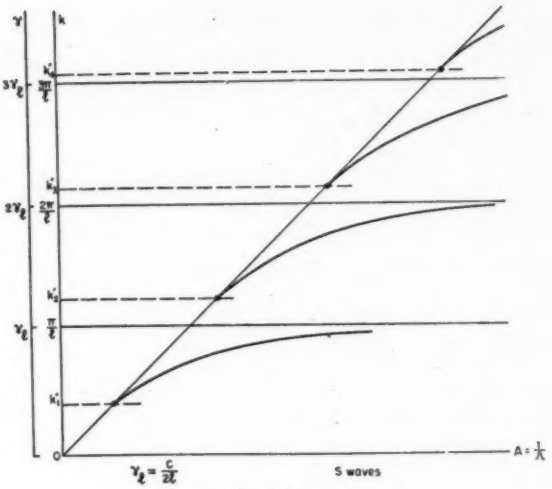


FIG. 23.

and (98) reduces to

$$-\frac{k'}{K}=t+\frac{k'}{K}, \text{ or } t=-2\frac{k'}{K} \text{ very small}; \quad (100)$$

the solution of which is

$$k'l=n\pi-2\frac{k'}{K}, \text{ where } n \text{ is an integer};$$

and for K infinite we find

$$k'=\frac{n\pi}{l}, \quad K=\infty. \quad (101)$$

The values k_1', k_2', k_3' obtained from (99) yield $K=0$ and $W=c$ according to (12). This corresponds to one side of the passing bands. The upper frequencies of the passing bands are found in (101) when K is infinite, making $W=0$ and

$$\omega=k'c=n(\pi c/l),$$
$$2/\lambda=n/l, \quad \nu=n\nu_l, \quad (102)$$

which corresponds to resonances in the open slots. These results are summarized in Figs. 22

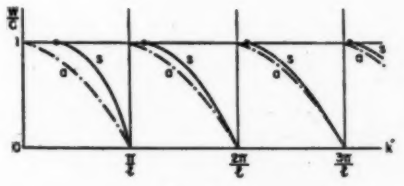


FIG. 24.

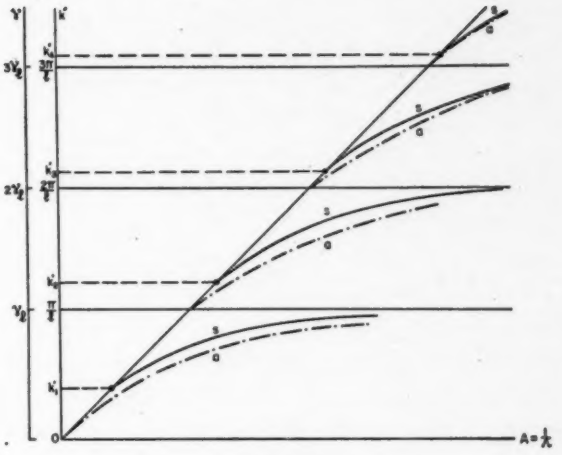


FIG. 25.

and 23. We now have to present a similar discussion for the *antisymmetrical waves*, where we use the formulae (74) in region III and (93) and (94) for the regions II and I. This changes condition (95) for field continuity on $y=b$, which now reads

$$\frac{H_x}{E_z}=-i\omega\epsilon_0\frac{\coth Kb}{K}=-i\omega\epsilon_0\frac{\tan\psi}{k'}, \quad (103)$$

and Eq. (96) is maintained. So we are left with a set of equations very similar to the previous one, but for the fact that

$$C=\coth Kb$$

now plays in the equations the same role as T previously did [compare (96) and (103)]. Hence the solution is similar to (98):

$$-C\frac{k'}{K}=\left(t+\frac{k'}{K}\right)\Big/\left(1-\frac{k'}{K}t\right), \quad t=\tan k'l. \quad (104)$$

Let us consider the limiting cases defined in (99) and (100):

$$K \text{ very small}, \quad C=1/Kb,$$
$$-k'/K^2b=-1/t, \quad (105)$$

and t is very small; hence

$$k'l=n\pi, \quad \nu=n\nu_l. \quad (105a)$$

Next we take K very large, and $C=1$, which gives the same limits as $T=1$; or again

$$k'l=n\pi, \quad \nu=n\nu_l. \quad (106)$$

The corresponding curves have been plotted in Figs. 24 and 25, where the symmetrical "s" and antisymmetrical "a" are shown together.

The following checks can be used to test the validity of these solutions:

A. *to increase b to infinity*. This, according to the diagram of Fig. 21, makes

$$k_1'=0, \quad k_2'=\frac{\pi}{l}, \quad k_3'=2\frac{\pi}{l}, \qquad (107)$$

and results in exact superposition of the "a" and "s" curves in Fig. 25. This makes the diagram of Fig. 25 identical with that of Fig. 18, except for a change in the ν scale. The length of the open slots is l in Fig. 16 and $2l$ in Fig. 20; hence ν_l in Fig. 25 corresponds to $\frac{1}{2}\nu_l$ in Fig. 18.

B. *to decrease b to zero*, thus making the two slot systems of Fig. 20 join together into the structure of Fig. 16. In doing this, we read from Fig. 21

$$k_1'=\frac{\pi}{2l}, \quad k_2'=3\frac{\pi}{2l}, \quad k_3'=5\frac{\pi}{2l}, \qquad (108)$$

and at the same time it can be proved that the horizontal asymptotes of the "a" curves in Fig. 25 are lowered to

$$\frac{\pi}{2l}, \quad 3\frac{\pi}{2l}, \quad 5\frac{\pi}{2l}.$$

Finally the diagram reduces exactly to that of Fig. 18.

XI. CYLINDRICAL STRUCTURE

Some general results can be obtained about wave propagation in a cylindrical structure as the one represented in Fig. 3. A detailed discussion would involve Bessel functions as explained in Section III, Eqs. (17) to (20). The general procedure would be very similar to that followed in Sections IV to VII for waves between two parallel plane structures, the only essential change being that exponentials e^{Ky} should be replaced by Bessel function $Z_0(iKr)$ if waves of cylindrical symmetry are considered ($n=0$), and exponentials e^{iKy} would be replaced by $Z_0(kr)$. The method is to start once again with a dominant wave in free space, and then adjust a dominant wave in the slots, in such a way as to insure continuity of E_z and H_θ along the cylindrical surface $r=b$. This would lead to a condition similar to (30). A great simplification would result from assumption (25)

$$d \ll l, \quad d \ll \Lambda,$$

and it is also convenient to consider the case when

$$b>l, \qquad (109)$$

since it corresponds to a structure where the dominant wave is very little perturbed along the axis of the cylinder. Here again, as in the plane structure of Section IV, the secondary free waves are rapidly attenuated and die out when $b-r$ is materially greater than d (see Eqs. (35) and (36)). Hence the dominant free wave would be practically unperturbed within a cylinder of radius $b-pd$ about the axis (p being of the order of a few units).

If we consider waves with no θ dependence, we obtain a system of successive passing bands very similar to the one described in Figs. 12 and 13. The position of the lower cut-off frequencies ν_1 and ν_4 of the passing bands is of course different, but it can be obtained very directly from the following physical discussion.

The lower cut-off frequency ν_1 between two parallel plane slot structures corresponds exactly to the lower cut-off between parallel planes at the distance h, and this is easy to understand. At this cut-off frequency the electric field is parallel to the planes everywhere, and runs in the two directions from $-\infty$ to $+\infty$. Hence this field is everywhere perpendicular to the fins and the presence of these fins does not alter the field at all. In the case of our cylindrical structure, we obtain exactly the same conditions and the same electric field distribution on the lower cut-off frequency for TM waves in a cylinder of radius $b+l$ when

$$k_{01}=\frac{2.405}{b+l}, \quad \nu_1=\frac{ck_{01}}{2\pi}=\frac{c}{\pi h}2.405, \qquad (110)$$

since $b+l=\frac{1}{2}h$. In a similar way, ν_4 corresponds to the lower cut-off of the next type of waves between parallel plates at a distance h and should be replaced by the corresponding

cut-off frequency for a cylinder of diameter h,

$$k_{02}=\frac{5.520}{b+l}, \quad \nu_4=\frac{ck_{02}}{2\pi}=\frac{c}{\pi h}5.520. \quad (111)$$

For parallel plane structures we had

$$\nu_4=3\nu_1,$$

while for the cylindrical structure we obtain

$$\nu_4=\frac{5.52}{2.405} \quad \nu_6=2.3\nu_1. \quad (112)$$

Otherwise, the band structure will be very similar to that of Figs. 12 and 13.

As for the upper cut-off frequencies, they were found at

$$\nu_3=\tfrac{1}{2}\nu_1 \quad \text{and} \quad \nu_6=\tfrac{3}{2}\nu_1 \quad (113)$$

on the resonance frequencies of the slots, but we noted that the approximation used in the present theory [especially condition (25a)] were no more reliable in the neighborhood of the upper cut-off frequency. In the cylindrical structure, there would be an additional correction as a result of the cylindrical shape of the slots but, at any rate, the order of magnitude of ν_3 and ν_6 should be correct and Eq. (113) can be considered as a rough first approximation.

Divergent Beam X-Ray Photography with Standard Diffraction Equipment

A. H. Geisler, J. K. Hill, and J. B. Newkirk*
General Electric Research Laboratory, Schenectady, New York

(Received May 4, 1948)

A new method for preparing divergent beam x-ray photographs of crystals is presented. This employs the usual Laue transmission camera and a collimated primary beam of x-rays. Fluorescent radiation originating either in the crystal sample or at a radiator in front of the crystal is used as the source of divergent x-rays. Some applications of the technique to determinations of orientation, lattice parameters, and crystal perfection are illustrated.

INTRODUCTION

THE general diffraction technique which will be described here has been the subject of a recent review.[1] Although the original observations which led to the development of the divergent beam technique were made over thirty years ago, we are indebted to Dr. Lonsdale for bringing the technique and its potentialities to the attention of those in this country who are interested in crystallography.[2] The authors' interest in this method of analysis evolved from incidental observations made during a study of single crystals of some solid solution alloys by the Laue technique. These observations demonstrated that the diffraction effects could be obtained under certain conditions with standard x-ray diffraction equipment.[3] A special x-ray tube designed to provide a highly divergent primary beam of x-rays[1,4,5] is not required. Suitable patterns can be made with the usual crystal analysis tubes—thus making the method available to the average crystallographer.

The technique is based on the Bragg diffraction which occurs when a conical beam of x-rays strikes a large single crystal specimen. If the angle of the cone is sufficiently large, those rays which strike a particular crystallographic plane at the required Bragg angle, θ, will be diffracted as cones. These may contribute to the transmitted background intensity of the pattern and appear as black lines. As a result, the background intensity will be diminished where these rays would have struck the film had they not been

* The last author now at Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

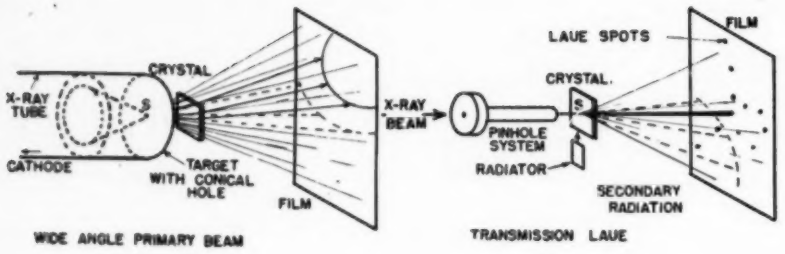
[1] K. Lonsdale, Phil. Trans. Roy. Soc. **240A**, 219 (1947).

[2] American Society for X-ray and Electron Diffraction, June 1947 Meeting, Ste. Marguerite Sta., P.Q.

[3] A. H. Geisler, J. K. Hill, and J. B. Newkirk, Phys. Rev. **72**, 983 (1947).

[4] H. Seemann: Ann. d. Physik **6**, 793 (1930); *ibid.* **7**, 633 (1930).

[5] B. Hess, Zeits. f. Krist. **97**, 197 (1937); *ibid.* **104**, 294 (1942).



a b

FIG. 1. Two methods for making divergent beam x-ray photographs; a—using a special tube that provides a primary source of divergent radiation and b—using a sealed-off tube and a secondary source of divergent radiation.

diffracted. The origin of the black and white lines thus produced will be more completely described later.

II. THE TECHNIQUE

The obvious requirement of the technique is a solid cone of x-rays of large apex angle. A primary source of high intensity (Fig. 1a) permits short exposures and large specimen to film distances but may not be available. Other methods[6–8] are not widely applicable. The method proposed by the authors (Fig. 1b) depends upon fluorescent secondary radiation as the source of the required divergent beam and thus requires conditions conducive to intense fluorescence. These are conditions which are generally avoided in the usual methods of crystal analysis in order to avoid extreme background blackening; they can easily be determined by consulting tables of absorption coefficients for various incident radiations.[9] A collimated beam of higher frequency x-rays is used to excite lower energy secondary radiation either at the specimen or at a thin film of a suitable material (radiator) placed immediately in front of the specimen when the specimen is a weakly fluorescing material.**,[10] The usual Laue spots are formed by the primary beam, while rays are diffracted from the divergent secondary beam to provide the black and white lines that are characteristic of this method of analysis.

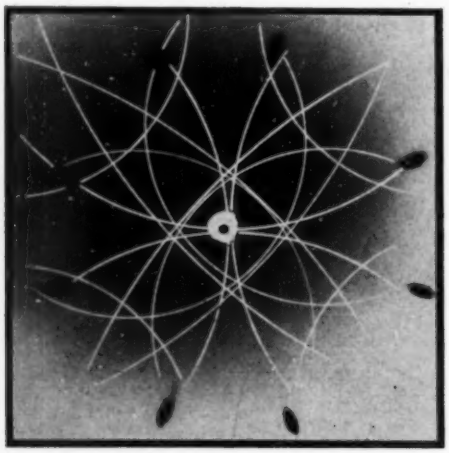


FIG. 2. Laue pattern of quenched Cunico (50 percent Cu, 29 percent Co, 21 percent Ni). Face-centered cubic crystal oriented with [001] parallel to incident beam of Cu radiation. White arcs originate in diffraction from background CoK_α secondary radiation.

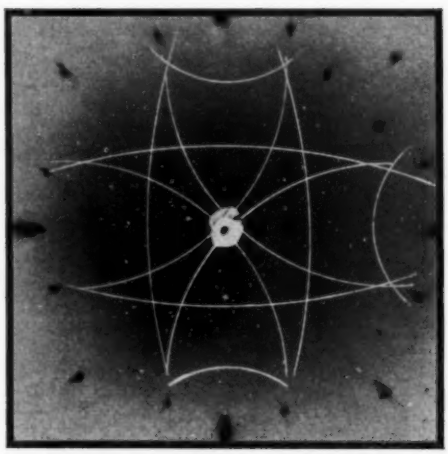


FIG. 3. Laue pattern of quenched Alnico 4 (55 percent Fe, 28 percent Ni, 12 percent Al, 5 percent Co). Body-centered cubic crystal oriented with [001] parallel to incident beam of Cu radiation. White arcs originate in diffraction from background FeK_α secondary radiation.

[6] W. Kossel and H. Voges, Ann. d. Physik **23**, 677 (1935).
[7] W. Kossel, Ann. d. Physik **25**, 512 (1936); *ibid.* **26**, 533 (1936).
[8] H. Voges, Ann. d. Physik **27**, 694 (1936).
[9] C. S. Barrett, *The Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 520.
** Borrmann (see reference 10) used a back reflection Laue method. Excellent patterns for Cu, Zn, Cr, Fe_3O_4, ZnS, $SrCO_3$, and KBr crystals were obtained with copper or molybdenum primary radiation; however, a provision for making this method applicable to the lighter elements or their compounds is not apparent.
[10] G. Borrmann, Ann. d. Physik **27**, 669 (1936).

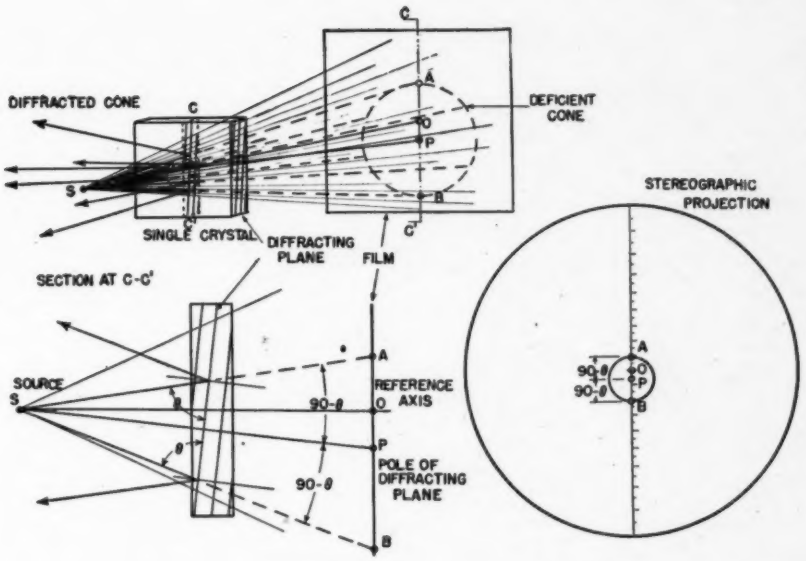


FIG. 4. Diagrams illustrating origin of white lines and method of plotting in stereographic projection.

Patterns made with primary copper radiation and crystals of alloys which contained cobalt or iron are shown by Figs. 2 and 3.*** Back reflection of appropriate rays from the highly divergent secondary CoK_α and FeK_α radiation results in the white lines through the background.

III. ORIGIN AND ANALYSIS OF THE PATTERNS

The drawing in Fig. 4 illustrates the origin of the white lines. The divergent beam originating at S is directed towards the single crystal.† Those rays which are incident to the diffracting plane at the Bragg angle, θ, are diffracted backwards as a cone and would register as a black circle on a film placed behind the source. The other rays pass through the specimen and blacken the film. The diffracted rays leave a cone in the transmitted radiation which is deficient in x-rays as shown by the dashed lines. This registers as a white circle, ellipse, or arc on the generally

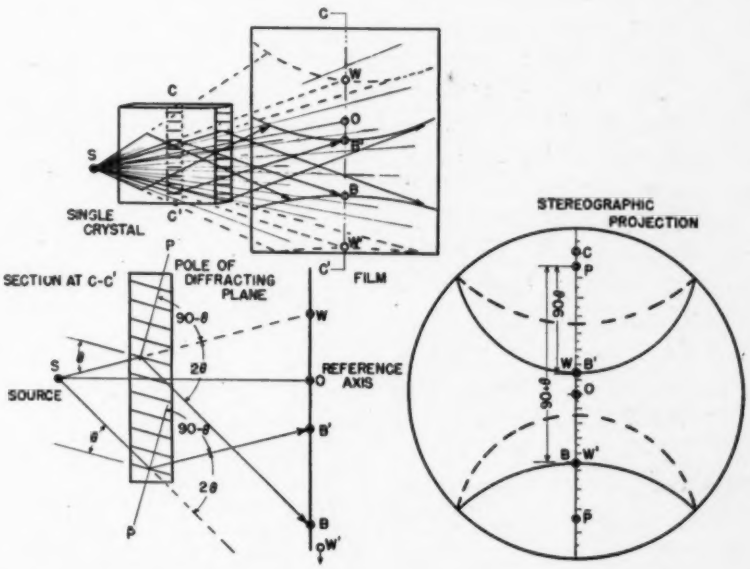


FIG. 5. Diagrams illustrating conditions under which both black and white lines appear in divergent beam photographs and method of plotting in stereographic projection.

*** To improve reproduction, the lines have been traced with ink.

† The source is shown at a distance from the specimen merely to simplify the drawing. Actually, it is within the specimen at the origin of the secondary radiation.

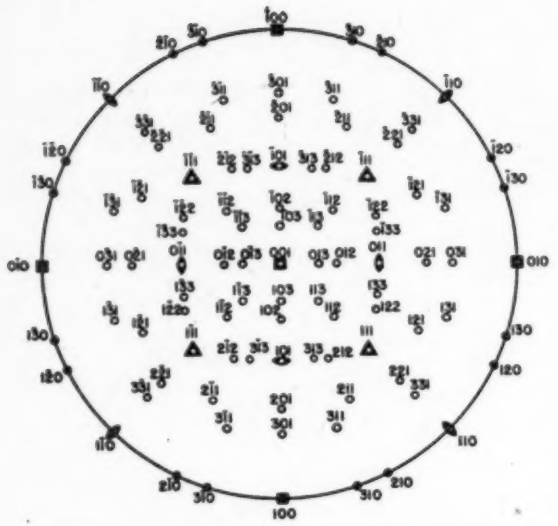


FIG. 6. Stereographic projection of a cubic crystal in a standard orientation.

blackened film. As shown in the cross section drawing, the deficient rays form a cone of half-vertex angle equal to $90-\theta$ with the pole of the reflecting plane, P, as axis. The stereographic projection[11] of the deficient cone (also diffracted cone) is then represented by a circle of radius $r=90-\theta$ about the pole of the diffracting plane. The center of the stereographic projection, O, was chosen as the horizontal of the drawing which may be the line perpendicular to the specimen or the direction of the primary collimated beam. A crystallographic axis of high symmetry is most conveniently chosen as center of the stereographic projection (see Figs. 7 and 8).

When the diffracting plane makes a small angle with the reference axis, and θ is small, both diffracted and deficient cones may be registered on the film as in Fig. 5. The separation between the black line, B', for diffraction from one side of the plane and the parallel white line, W, for the other side of the plane depends upon the distance from the source to the point of diffraction along a direction normal to the photographic film.[1] In the stereographic projection the cones for the two superimpose, since angle alone is considered. These are true circles, but their center, C, is not the projected axis of the cone, P. This is a characteristic of the stereographic projection. The portions of the circles which are in the rear hemisphere and which correspond to the back reflected pattern have been indicated by the broken lines in the stereographic projection in Fig. 5.

The lines are identified by calculating values of $90-\theta$ for the material and radiation, using Bragg's law in the form (for cubic crystals):

$$r=90-\theta=\cos^{-1}\lambda(h^2+k^2+l^2/2a_0)^{\frac{1}{2}}. \qquad (1)$$

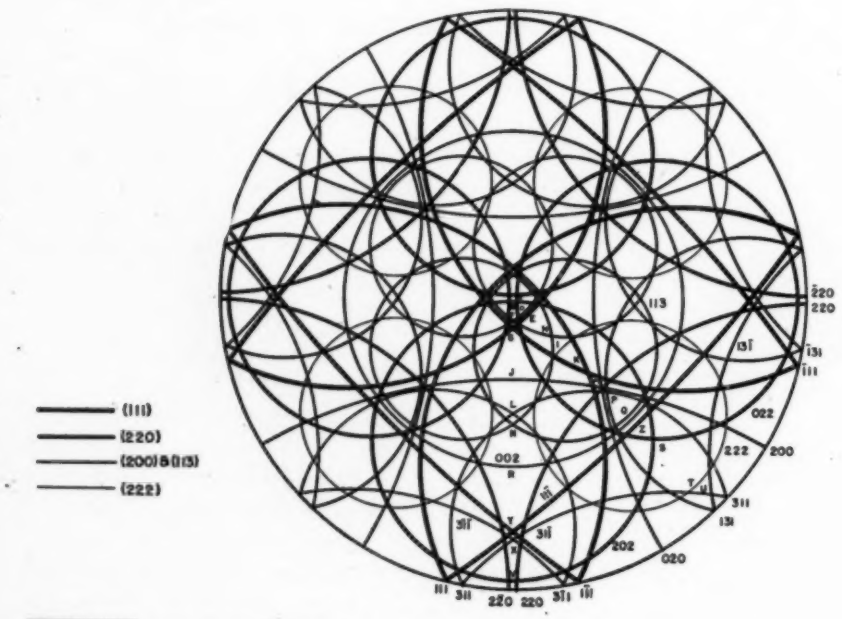


FIG. 7. Stereographic projection of pattern calculated for Cunico. Face-centered cubic crystal with $a_0=3.53$A in the orientation of Figs. 2 and 6. For $CoK_{\alpha 1}$ radiation.

[11] See reference 9, pp. 25–43.

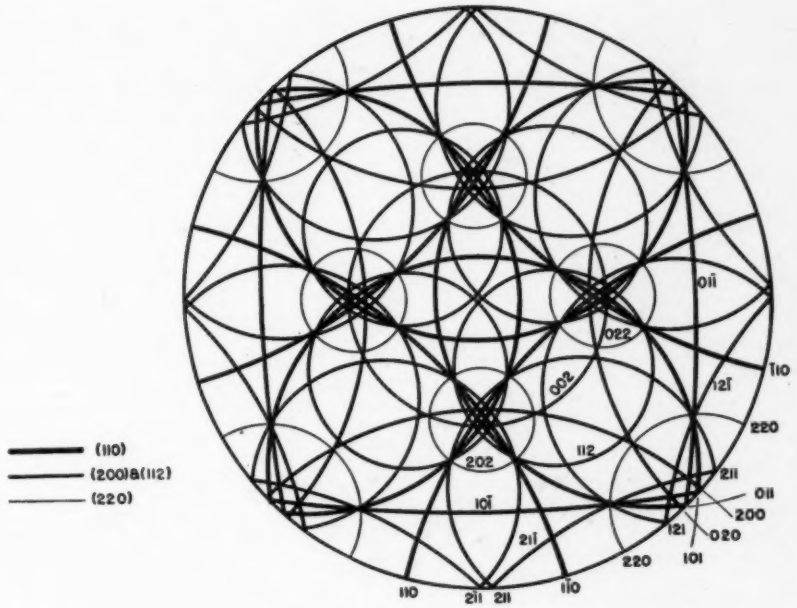


FIG. 8. Stereographic projection of patterns calculated for α-iron. Body-centered cubic crystal with $a_0 = 2.861$ in the orientation of Figs. 3 and 6. For $FeK_{\alpha 1}$ radiation.

These values are used to draw the projections of the cones about the corresponding poles of the crystal in a standard position, such as shown by Fig. 6. Projections predicted for Cunico with CoK_α radiation and for Fe with FeK_α radiation are shown by Figs. 7 and 8. The indices of cones for some of the planes are listed on the drawings. The others may be obtained by reference to Fig. 6. Direct comparison of the patterns with the stereographic projections is possible. In Fig. 2 and Fig. 7 the most intense features are the four (111) type lines which intersect at the 45° positions (at central portion of Fig. 7). Next, there are the (202) type arcs which pass near the incident beam and are nearly tangent at the 0(90°) positions. Finally, there are the less intense (113) type lines which are almost tangent to the (111) lines at the 45° positions. The lattice parameter of alnico is little different from that of αFe—thus permitting the comparison of Fig. 3 with Fig. 8. This will identify the lines in Fig. 2 as the more intense (101) and (112) types and the faintly visible (202) type.

IV. SOME RESULTS WITH AUXILIARY RADIATORS

A few experiments were made to survey briefly the possibilities of using divergent secondary radiation excited at an external source for the study of crystals which themselves are not good radiators. The results are encouraging. Combinations of Cu incident radiation and iron, iron oxide, and cobalt radiators were investigated. The latter materials were secured as -325-mesh powder, which was mounted on Cellophane tape. This was placed directly in front of the crystals being investigated, as illustrated by Fig. 1. Some of the patterns obtained for rocksalt and aluminum are reproduced in Fig. 9. The deficient and diffraction conics, while most prominent in the central area of scattering, extended further from the central region in some instances, notable in Fig. 9a. By changing the orientation of the specimen, any region of the pattern which might be of special interest can be moved into a more intense region of the divergent beam. The survey was by no means exhaustive in either determining the most suitable source of secondary radiation†† or in finding the optimum values of specimen thickness and radiator thickness.

V. APPLICATIONS

The applications of the divergent beam technique to studies of crystal perfection and to precision measurements of lattice constants and wave-lengths have been fully reviewed and illustrated in regard to diamond, ice, and some organic compounds.[1] A few applications of these types to metals can now be illustrated. All of the informa-

†† Other combinations are suggested by the ** footnote on page 1042.

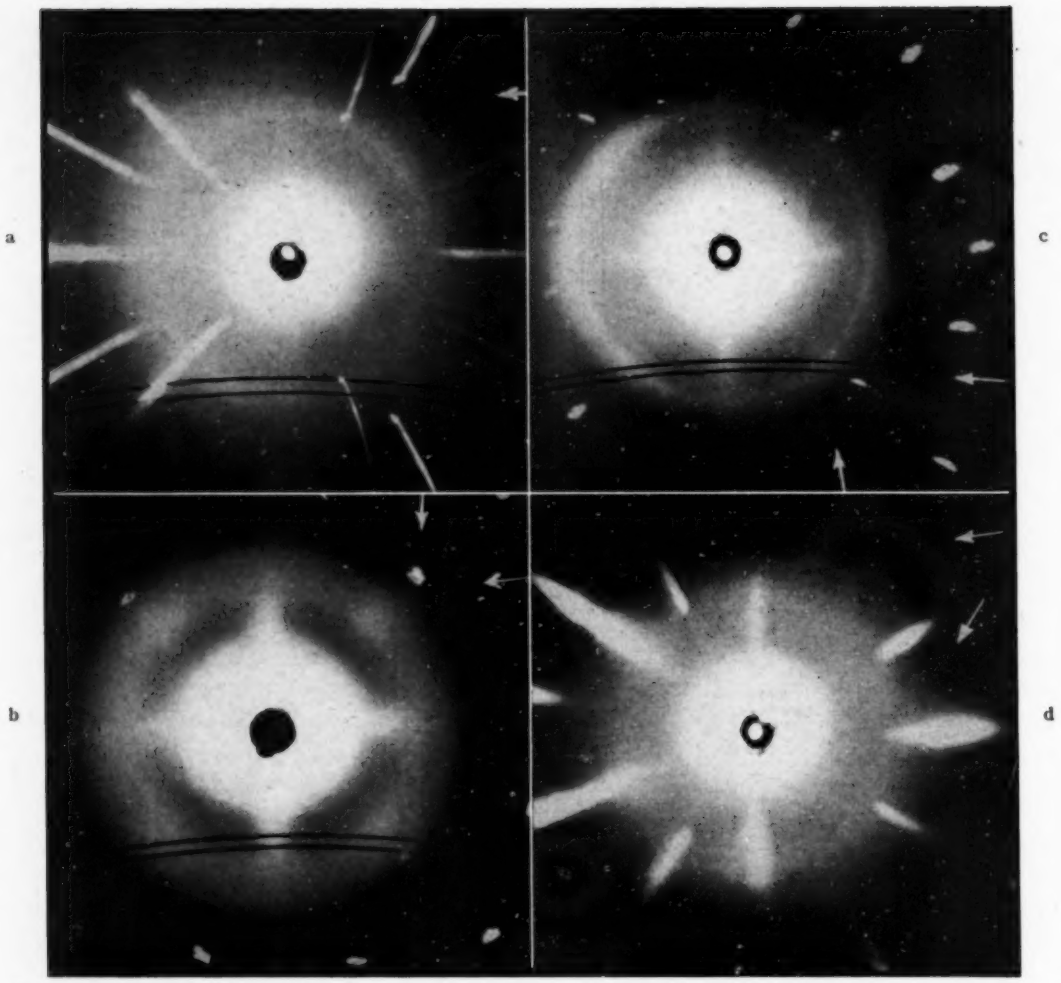


FIG. 9. Laue patterns for crystals that do not provide intense secondary radiation. Weak black and white lines are due to diffraction from secondary radiation excited in an auxiliary radiator placed in front of the single crystal specimen. Cu primary radiation. a is distorted rocksalt crystal with Fe auxiliary radiator; b is rocksalt crystal with FeO axuiliary radiator; c is rocksalt crystal with Co auxiliary radiator; d is distorted aluminum crystal with Co auxiliary radiator.

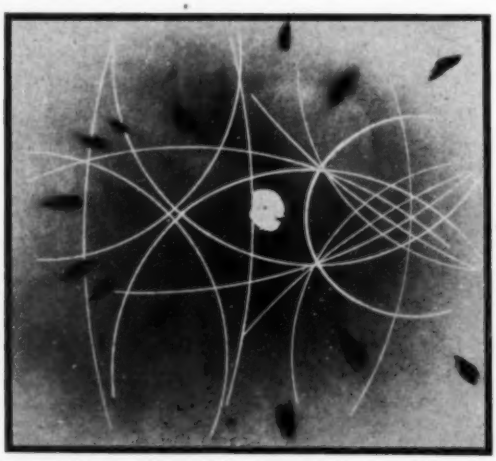


FIG. 10. Pattern for Alnico crystal inclined about [100] 20° from [001] parallel to the incident beam of copper radiation.

tion is obtained by inspection of the patterns—the geometry, width and intensity of the white lines. Calculations are not made on the patterns, and the specimen-to-film distance, film orientation, and specimen orientation need not be accurately known; however, the crystal structure, the approximate lattice parameter, and the wave-length of the diffracting, divergent radiation must be known.††† These are used to calculate the cone angles according to Eq. (1) which are then used to plot the expected pattern, such as Fig. 8, in order that the lines can be identified.

††† For precise measurements of wave-length, the approximate wave-length and precise lattice constant of the diffracting crystal are required.

The orientation of the crystal is immediately obvious upon comparing the observed pattern with the stereographic plot of the expected pattern. Prominent configurations about elements of symmetry are used. For example, the three lens-shaped areas formed by portions of arcs within the (022) circles of Fig. 8 are immediately recognizable at the right in the pattern reproduced in Fig. 10. The fourfold [001] configuration of Fig. 3 is left of the center of Fig. 10. If the primary direct beam is used as reference axis, its position can be determined relative to lines in the pattern and in the stereographic plot. Inspection of the array of Laue spots does not provide information about the crystal orientation so clearly.

The measurement of lattice parameters from the divergent beam patterns is based upon the location of points of intersection of deficient cones relative to each other or to lines very close to the intersection (features such as are indicated by letters in Fig. 7). With small changes in lattice parameter, the relative positions of these features change, since the angles of the cones depend upon lattice parameter. The requirements are therefore: (a) a portion of the pattern in which there are very near coincidences of intersections and lines and (b) a plot showing how these features change with lattice parameter, for comparison. Such a plot is illustrated by Fig. 11. In this, the positions of the designated points along the 0(90) and 45° azimuths of Fig. 7 in terms of angular distances from the (001) pole are plotted against lattice parameter covering the range of interest. By spherical trigonometry these distances can be determined from the radii of the cones (Eq. (1)) and the angular positions of the axis of the cones (poles of diffracting planes in the stereographic projection) relative to each other and to the (001) pole.‡ A table of the angles between poles for

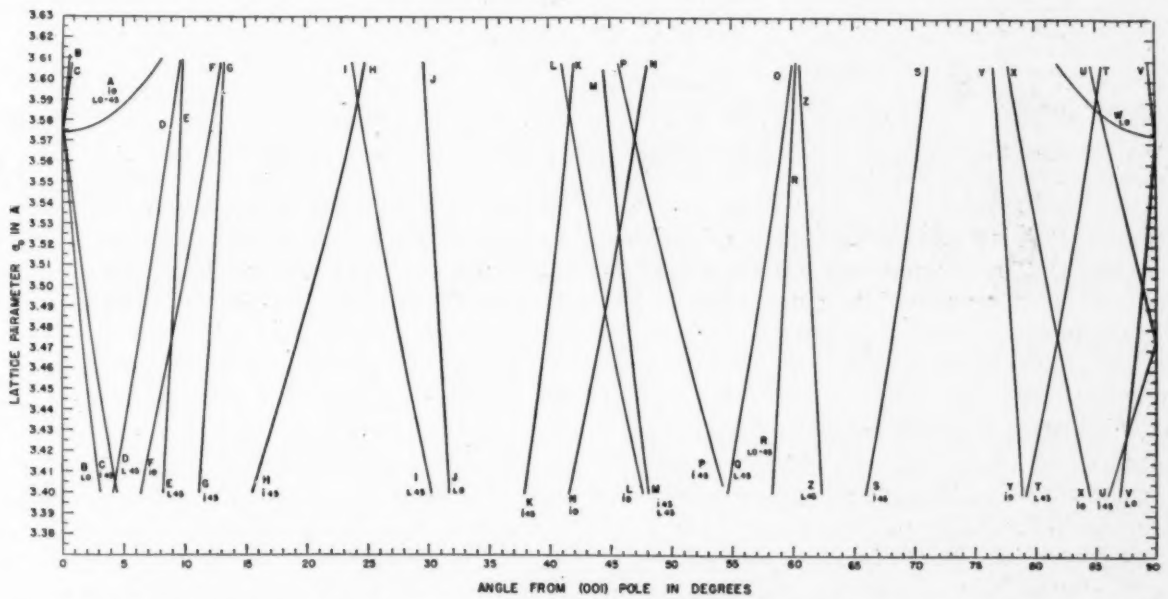


FIG. 11. Locii of critical points of Fig. 7 as a function of lattice parameter for $CoK_{\alpha 1}$ radiation. The designations "*i*0" and "*i*45" refer to line intersections at azimuth angles of 0° and 45° in Fig. 7, whereas "*L*0" and "*L*45" refer to lines tangent to radii at the two azimuth angles.

‡ The cube pole (001) is merely used as the reference axis. Actual measurements on the observed pattern are not made. Right spherical triangles are solved to give expressions of the following form for the intersections at 0, 54, or 90° aximuths:

$$ox' = \cos^{-1}(\sin\theta/(\cos pp'/2)] \pm \cos^{-1}[\cos op/(\cos pp'/2)], \tag{2}$$

where ox' is the desired angle, θ the Bragg angle given by expression (1), pp' the angle between the poles of the two planes of the same family which produce the intersecting cones, and op the angle from [001] to the pole of either plane. The nearest and farthest distances from [001] to the various cones are:

$$ox'' = op \pm (90 - \theta) \tag{3}$$

along a line from [001] to the pole of the cone axis.

TABLE I. Location of critical points in patterns of cubic materials as function of wave-length and lattice parameter.

Point*	Description	Azimuth	Distance from (001) pole
A	Any point on 004	All	$OA=\cos^{-1} 2\,\lambda/a$
A'	Intersection 022 and $0\bar{2}2$	0	$OA=\cos^{-1} 2\,\lambda/a$
B	Nearest point on 202	0	$OB=45°-\cos^{-1} 1.4142\,\lambda/a$
C	Nearest intersection 202 and 022	45	$OC=35°\,16'-\cos^{-1} 1.6330\,\lambda/a$
D	Nearest point on $\bar{1}\bar{1}3$	45	$OD=\cos^{-1} 1.6583\,\lambda/a-25°\,14'$
E	Nearest point on $\bar{1}\bar{1}1$	45	$OE=\cos^{-1} 0.8660\,\lambda/a-54°\,44'$
F	Nearest intersection $\bar{1}13$ and $\bar{1}\bar{1}3$	0	$OF=\cos^{-1} 1.7392\,\lambda/a-17°\,33'$
G	Nearest intersection $\bar{1}11$ and $\bar{1}\bar{1}1$	0	$OG=\cos^{-1} 1.0607\,\lambda/a-45°$
H	Intersection $\bar{1}13$ and $1\bar{1}3$	45	$OH=\cos^{-1} 1.8333\,\lambda/a$
I	Nearest point on 222	45	$OI=54°\,44'-\cos^{-1} 1.7321\,\lambda/a$
J	Nearest point on 200	0	$OJ=\sin^{-1}\lambda/a$
K	Intersection $1\bar{1}1$ and $\bar{1}11$	45	$OK=\cos^{-1} 1.5\,\lambda/a$
L	Intersection 311 and $3\bar{1}1$	0	$OL=71°\,34'-\cos^{-1} 1.7392\,\lambda/a$
M	Nearest point on 220	45	$OM=\sin^{-1} 1.4142\,\lambda/a$
M	Intersection 200 and 020	45	$OM=\sin^{-1} 1.4142\,\lambda/a$
N	Farthest intersection 113 and $1\bar{1}3$	0	$ON=\cos^{-1} 1.7392\,\lambda/a+17°\,33'$
P	Intersection 311 and 131	45	$OP=70°\,32'-\cos^{-1} 1.835\,\lambda/a$
Q	Farthest point on 113	45	$OQ=\cos^{-1} 1.6583\,\lambda/a+25°\,14'$
R	Any point on 002	All	$OR=\cos^{-1}\lambda/a$
S	Farthest intersection 202 and 022	45	$OS=35°\,16'+\cos^{-1} 1.6330\,\lambda/a$
T	Farthest point on 222	45	$OT=54°\,44'+\cos^{-1} 1.7321\,\lambda/a$
U	Intersection $31\bar{1}$ and $13\bar{1}$	45	$OU=70°\,32'+\cos^{-1} 1.8335\,\lambda/a$
V	Farthest point on 202	0	$OV=45°+\cos^{-1} 1.4142\,\lambda/a$
W	Nearest point on 040	0	$OW=\sin^{-1} 2\,\lambda/a$
X	Intersection $31\bar{1}$ and $3\bar{1}\bar{1}$	0	$OX=108°\,26'-\cos^{-1} 1.7392\,\lambda/a$
Y	Intersection $11\bar{1}$ and $1\bar{1}\bar{1}$	0	$OY=135°-\cos^{-1} 1.0607\,\lambda/a$
Z	Nearest point on $11\bar{1}$	45	$OZ=125°\,16'-\cos^{-1} 0.8660\,\lambda/a$

* See Fig. 7.

cubic crystals is readily available (see reference 9, pp. 25–43). The expressions for the angular distances from the center of the pattern of Fig. 7 to the variously lettered critical points are listed in Table I. These are applicable for the face-centered reflections of any cubic material and wave-length of divergent radiation. Shorter wave-lengths and other cubic arrangements involve some of these in addition to others: iron requires expressions for (110) and (112) intersections also.

The extensive plot of Fig. 11 is not required for precise lattice parameter determinations but only a specific area suitable for the particular approximate lattice constant. For example, curves H and I are suitable for giving precise lattice constants when the approximate a_0 is between 3.585A and 3.595A and CoK_α radiation is used. Curves C and D are suitable for lattice parameters in the range of 3.40A to 3.41A, curves P and Q for 3.39A to 3.40A, and curves T and U for 3.59A to 3.60A. For high precision, the lattice parameter and angle scales within these ranges are expanded, and parallel curves are calculated for both α_1 and α_2 wave-lengths. Thus, resolution of the K_α doublet in the pattern is desirable. On comparison of the particular enlarged‡‡ portion of the observed pattern with the calculated curves, the precise lattice parameter for which the arrangement of curves satisfies the observed pattern can be assigned to the crystal. This method has been used to measure lattice parameters of diamonds to the *fifth* decimal place in angstroms.[1]

In determining the lattice parameters of the phases in the alloy Cunico, an accuracy to the third decimal place must suffice, for the lines are relatively diffuse and the $\alpha_1-\alpha_2$ doublet is not resolved. The pattern for the homogeneous sample reproduced in Fig. 12a shows that the four (202) and (022) type curves intersect at a common point—the (001) pole. Thus, the distance OB is zero, and according to Fig. 11 the lattice parameter is about 3.574A. If the doublets

‡‡ Enlarged merely for better resolution on visual examination.

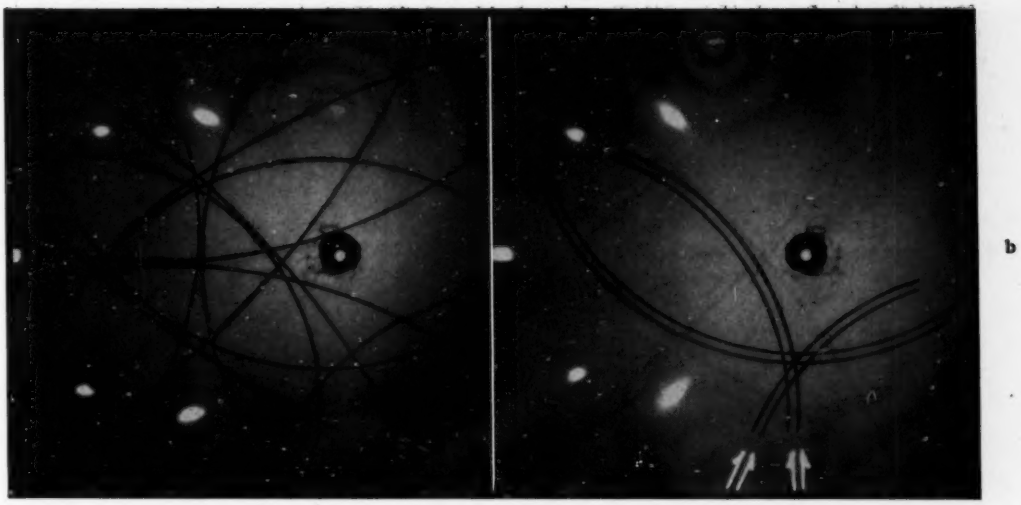


FIG. 12. Patterns for Cunico crystals inclined about [100] 20° from [001] parallel to the incident beam of copper radiation. a—as quenched from 1100°C—one phase. b—aged 42 hours at 700°C—two face-centered cubic decomposition products of slightly different lattice parameters.

had been resolved, improved accuracy could be obtained by recording this portion of the pattern on a fine grained emulsion and enlarging it photographically several diameters. Comparison with curves *B* and *C* plotted for CoK_{α_1} and K_{α_2} for a_0 between 3.57 and 3.58 would permit greater accuracy in assigning the lattice parameter.

When this alloy is reheated for a sufficient length of time it decomposes into two face-centered cubic phases of slightly different lattice parameters. Some of the lines are sharply resolved into pairs, as in Fig. 12b. Doubled (222) and (113) lines intersecting at *H* (Fig. 7) are evident. Similar doubling of the CoK_α (111) lines is not resolved. The outside (111) lines apparent in Figs. 2, 12a, and 12b must be attributed to scattered CuK_α radiation, since it also appears in patterns of pure copper. The near coincidence of intersection *H* with line *I* for the phase with the larger lattice parameter in Fig. 12b suggests that the lattice parameter is about 3.590A. The lattice parameter of the other phase apparently is less than that of the homogeneous alloy (Fig. 12a) because of the larger separation between points *H* and *I* in the pattern. A suitable coincidence of white lines was not observed with divergent cobalt (secondary) radiation.

Applications of the divergent beam technique to studies of crystal perfection are based on the line width in a qualitative manner. The line sharpness increases with increased crystal perfection. With an ideally perfect crystal, a pattern is not obtained because of insufficient line width for detection. On the other hand, the pattern of a grossly imperfect crystal is too diffuse for detection. Mention has been made of the diffuseness of the lines for Cunico. This is not characteristic of the method of production of divergent rays used by the authors, since quite sharp lines have been observed for rocksalt (Fig. 9a and b). The line diffuseness can best be attributed to the structure of the crystal specimens. Metals have low degrees of perfection, and the lines are diffuse, as shown by the patterns for as-quenched Cunico and for aluminum (Fig. 9d). The patterns for distorted crystals in Fig. 9 permit a comparison of the effects of distortion on the divergent beam lines and on the Laue spots. The Laue spots for rocksalt in Fig. 9a are spread into long streaks, while the black and white lines are relatively sharp. On the other hand, the Laue spots are only somewhat elongated in the pattern for aluminum (Fig. 9d), but the lines are rather diffuse. In contrast with the Laue pattern, the divergent beam pattern is more dependent on lattice parameter and less on orientation of individual blocks within a deformed crystal. The two methods of analysis combined should provide useful information concerning the plastic deformation process.

An Ion Microscope with a Transverse Magnetic Field

NOBUJI SASAKI
Institute of Chemistry, Department of Science, University of Kyoto, Kyoto, Japan
(Received April 26, 1948)

A brief discussion of electron and ion microscopes is presented followed by a description of a low magnification, electrostatic ion microscope. This instrument is of the emission type wherein the source of positive ions is imaged. If more than one species of ion is emitted, the images corresponding to the various atom weights may be separated by means of a transverse magnetic field and the ratio of atomic weights calculated from the displacements. The possible uses of such an instrument are pointed out.

THE electron microscope with a thermoelectrically or photoelectrically emitting cathode is in resolution and magnification very inferior to the well developed electron microscope with transmitted electrons. It has, however, its own merit, that it permits the study of the cathode surface in a way otherwise hardly possible. Thus much valuable knowledge has been obtained of the high temperature behavior of atoms in metal and semiconductors as well as adsorbed atoms on the surface. But the knowledge is indirect and incomplete in the sense that the electrons, being common to all substances, cannot differentiate the behavior of different elements present in a simple manner.

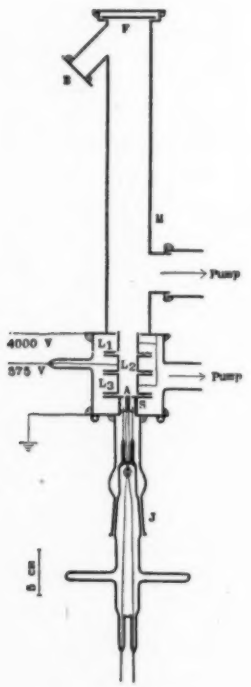


FIG. 1.

The atoms on the surface can be emitted as positive ions under suitable conditions. If these free ions, just like free electrons, could be used as image-forming rays in the microscopy, much more direct knowledge of the surface could be obtained. This is because the ions with masses characteristic of different elements can be made to give rise to separate ion images proper to each element or, precisely, to each isotope present in the ion beams.

The possibility in principle of the ion microscopy with a lens of electrostatic or magnetic type was suggested by Knoll and Ruska.[1] An ion microscope without a lens can also be constructed similar to the electron microscope of Johnson[2] and Shockley. A lens for heavy, charged particles such as ions must be of the electrostatic type for the obvious practical reason[3] that sufficiently powerful magnetic lenses cannot be constructed. The mass of the particle being indifferent for this type of lens, an electron microscope with an electrostatic lens can at once be converted into an ion microscope by reversing the applied potential.

In Fig. 1 showing our ion microscope, the anode *A* is an iron cylinder 2 mm in diameter, which can be heated by a tungsten filament inserted in its boring. Its upper surface, polished and covered with a barium-strontium mixture, gives a sharp electron image on the fluorescent

[1] Knoll and Ruska, Zeits. f. Physik **78**, 336 (1932).
[2] Saski and Mitani, Proc. Imp. Acad. Tokyo **17**, 36 (1941); Johnson and Shockley, Phys. Rev. **49**, 436 (1936).
[3] Sasaki and Mitani, Proc. Imp. Acad. Tokyo **19**, 156 (1943).

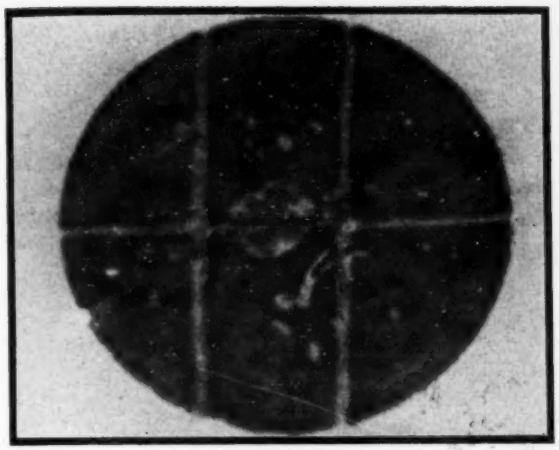


FIG. 2.

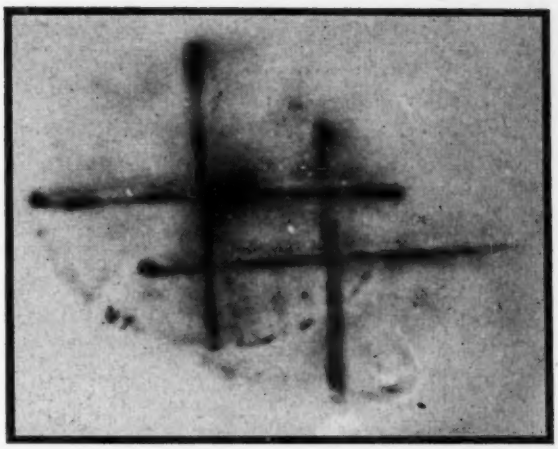


FIG. 3.

screen F (Fig. 2, a photograph taken through the window B) with L_1 and L_3 fixed at 4000 volts and L_2 adjusted at 575 volts. The necessary adjustment of the position of A relative to the guard ring S may be done from outside by a mechanism through the ground joint J. The magnification is about six.

To obtain an ion image an intimate mixture of potassium chloride and tungsten powder is rubbed into the razor scratches in the form of a cross and the anode is heated at 800°C with the same but reversed potential. The ion image thus formed on the zinc silicate layer on a glass plate is not so sharp as the electron image. Koch and Walcher,[4] who were the first to work with an ion microscope, could not obtain any ion image of an emitting anode. They attribute the failure to an intense space charge formed in front of the highly emissive anode surface. However, we would rather seek the reason in the charge accumulated on the fluorescent layer caused by the high resistance of the surface. Actually, the sharpness is improved if a thin fluorescent layer is deposited on a metallic plate or if the screen is bombarded with slow electrons for the purpose of neutralization of the ions accumulating there, though the brightness of the image is weakened thereby. With a thick layer a regular, pulsating intensity is observed. The best result was obtained with a fine metal wire gauze filled with a fluorescent powder.[5] Figure 3 shows a photograph of a potassium ion image (right cross) which can be shifted without any appreciable distortion and blurring when a magnetic pole M is placed in the position shown in Fig. 1.

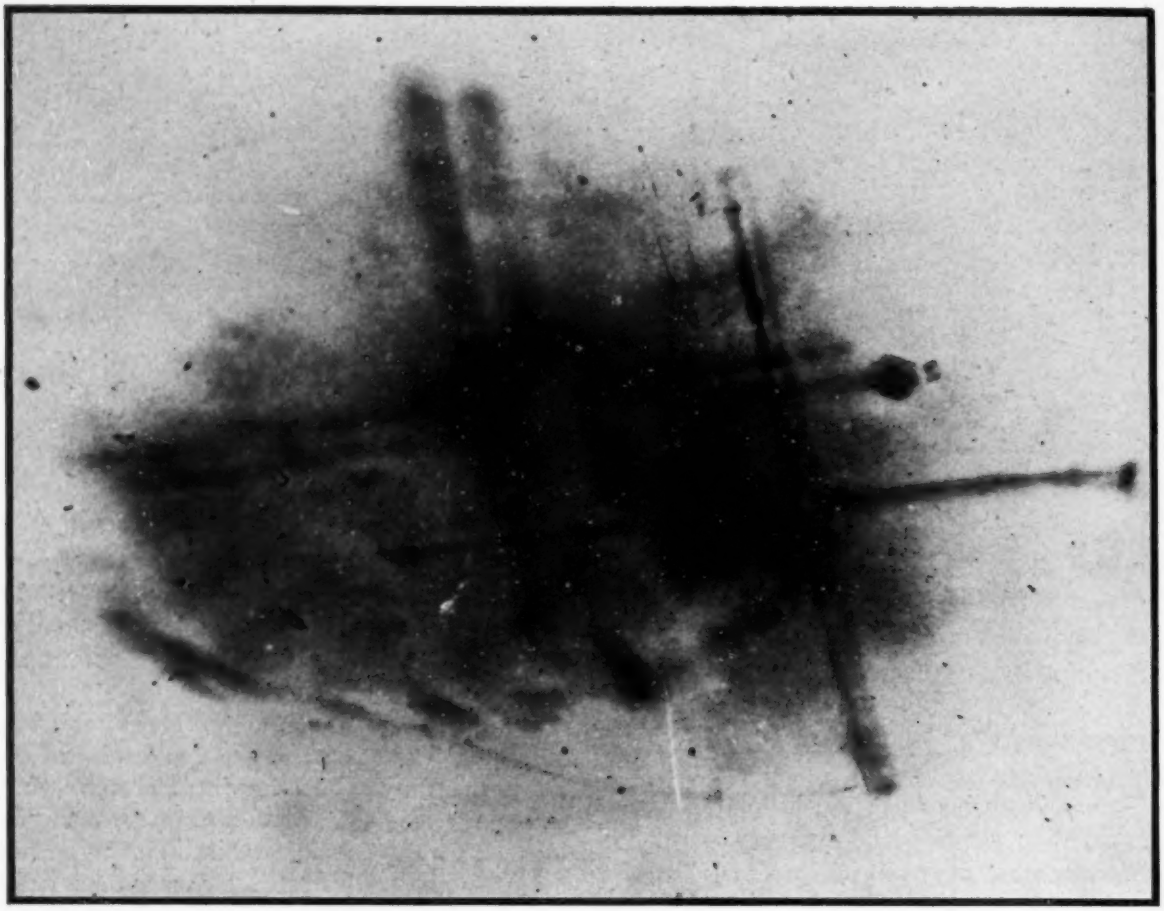
The admixture of sodium chloride to potassium chloride does not change the appearance of the ion image, but on applying the transverse magnetic field the single composite image separates into the component images, each making a shift inversely proportional to the square root of the mass number of the ion. In Fig. 4 the crosses correspond, from right to left, to the $Na^+ + K^+$ composite image, the K^+ image, and the Na^+ image. The actual shifts measured on the photograph are nearly in the ratio $(39)^{\frac{1}{2}}:(23)^{\frac{1}{2}}$. Conversely, the amount of the displacement made by an ion image can tell unambiguously the mass number and consequently the species of chemical element which forms the image (the occurrence of isotopes is very rare).

Thus the ion microscope with a transverse magnetic field can give not only a magnified image of surface structures, as do other types of microscopes, but also information about the kind of the chemical elements present and their surface distributions. It plays, therefore, a double role of microscope and mass spectrograph.

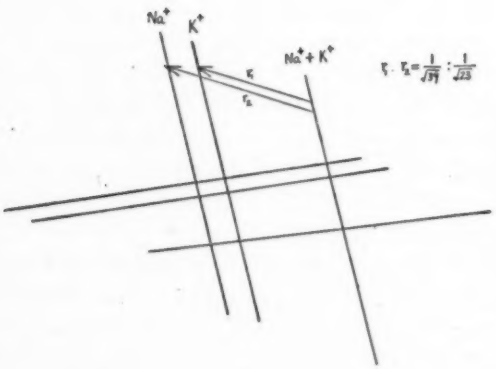
Our provisory ion microscope, though of very low magnification, can still be used to obtain

[4] Koch and Walcher, Zeits. f. Physik **97**, 131 (1935).

[5] A piece of this fluorescent screen called "amiscope" was kindly given me by Mr. Takeo Seki, engineer of the Communication Ministry, who invented a special method of preparation.



a



b

FIGS. 4a and 4b. The original ion microscopic image is a single cross consisting of Na^+ and K^+. It is shifted by a transverse magnetic field in the direction of arrows to be resolved into two component images, a Na^+ cross and a K^+ cross. The lengths of the arrows drawn in proportion to inverse square roots of the atomic weight of Na (23) and of the main isotope (39) of K well fit the actual displacements (Nobuji Sasaki and Kazuo Mitani, "Ion micro scope with a transverse magnetic field," Proc. Imp. Acad. Tokyo **19**, 156 (1943)).

unique information concerning surface phenomena. A direct observation through the window enables one to see how the emission state of the anode surface covered with alkali salts changes with the time. The whole surface from which any excess potassium chloride has been wiped off carefully becomes, on the first heating, equally ion emissive, while the cross alone remains dark. As the evaporation of potassium atoms progresses, each center of the four quadrants begins to give rise to a small dark spot, which expands slowly on all sides, the boundary always remain-

ing strikingly sharp. The bright areas of the quadrants recede continuously until only the narrow portions lying along both sides of the scratches remain still visible. Finally, the cross alone becomes visible and remains so for a comparatively long time. The behavior of the potassium atoms will not be changed by the addition of sodium chloride to potassium chloride. The ionization potential of the potassium atom is lower than that of the sodium atom. Accordingly, if the temperature of the anode is raised slowly, the potassium image appears at first. If the anode is heated rapidly to a temperature sufficiently high for sodium ion emission, no sodium image appears so long as a large number of potassium atoms are present in the mixture. There is observed at first the potassium ion emission alone, which decays gradually and a simultaneous faint emission of sodium ions begins. The latter gains in intensity until it replaces entirely the potassium emission. The phenomenon could be accounted for by a process represented by $Na^+ + K = Na + K^+$.

If the ion beam is replaced by a light beam emitted by the surface atoms and the electrostatic lens and the magnetic field by an optical lens and a prism, the ion microscope is transformed into its optical analog which can be called an optical spectral microscope.[6]

My deep thanks are due Mr. Kazuo Mitani who carried out the present experiments.

[6] Sasaki, Proc. Imp. Acad. Tokyo **20**, 368 (1944).

Growth of Charged Particles in Clouds*

WULF KUNKEL
Department of Physics, University of California, Berkeley, California
(Received May 20, 1948)

In order to calculate the average growth of a charged particle in a dust cloud caused by electrostatic attraction, the following assumptions were made: the particles obey Stokes' law of motion, their velocities are small so that their motion can be considered to be in constant equilibrium with the forces acting, and the forces are effective at short ranges only or the cloud density is small enough so that not more than two particles have to be considered interacting at a time. In addition to the Coulomb force, there will appear an attraction between particles resulting from induced dipoles, but it was found that in general the effect of this force can be neglected. Straight forward integration of the equation of relative motion of two charged dust particles yields their effective cross section for aggregation from which in turn we can determine the rate of growth if the particle size and charge distribution in the cloud is known. Assuming a highly simplified distribution, one arrives at the result that aggregation is negligible if the cloud density is less than a million particles per cm^3 or if the average charge of one sign is well below a thousand electrons per particle.

INVESTIGATIONS of charge and particle size in dust clouds studied by means of the Hopper and Laby method for the determination of the electronic charge made it imperative to compute the conditions for aggregation of the particles in the cloud as it settled. Since this problem may be of interest to other investigators, a brief summary of the analysis is herewith presented.

When small dust particles are whirled up or fine droplets of a liquid are sprayed to form a cloud, the individual particles will in general be found to carry charges of varying magnitude and both signs. If the cloud is allowed to settle under gravity, opposite charges will tend to draw the particles together forming aggregates thus affecting their average size. The following rough calculations enable us to estimate the conditions of cloud density and average charge that will result in appreciable particle growth.

We assume perfect spheres that obey Stokes'

* Acknowledgment is made to the Office of Naval Research for assistance under Contract No. N7ONR295, T.O. IX, which made possible the obtaining of some of this data.

law of motion (size 1 to 20 microns diameter). On touching, two spheres shall combine to form one sphere of volume equal to the sum of the two original volumes (liquid droplets). Except for the electrostatic forces between them, the droplets are subject to gravity only, and the forces of interaction are short ranged enough to affect only two neighboring particles at a time.

In addition to the Coulomb force between two charges, there will be an attraction caused by induced dipoles appearing in the particles. Assuming further that the velocities in question are small enough to be constantly in equilibrium with the forces (given by Stokes' law) we can write down for the relative velocity of two particles as a result of electrostatic forces

$$V_r = C_1/r^2 + C_2/r^5, \tag{1}$$

where

$$C_1 = \frac{q_1 q_2 (a_1+a_2)}{6\pi\eta a_1 a_2},$$

$$C_2 = -\frac{2(K-1)(a_1^3 q_2^2 + a_2^3 q_1^2)(a_1+a_2)}{(K+2)6\pi\eta a_1 a_2}.$$

Quantities referring to the two particles are distinguished by the subscripts 1 and 2; q is the charge; C_1 is negative if q_1 and q_2 are of opposite sign so that we have a Coulomb attraction. a is the radius of a particle, η the viscosity of the medium, and K the dielectric constant of the particles. C_2 is always negative. The expression is, however, only approximate, C_2 increasing slightly as r decreases.

In general now $a_1 \neq a_2$ so that the particles will also have a relative velocity caused by gravity.

$$V_0 = V_1 - V_2 = \frac{2g\rho}{9\eta}(a_1^2 - a_2^2).$$

(ρ = density of particles.) (2)

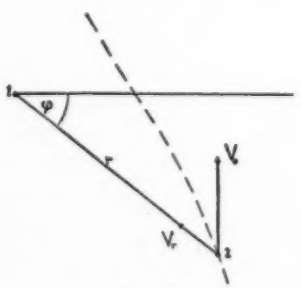


FIG. 1.

Consider particle 1 fixed at the origin of a system of polar coordinates. Then, since V_r is in the direction of the radius vector and V_0 is vertical (see Fig. 1),

$$\frac{dr}{dt} = V_r + V_0 \sin\phi, \quad r\frac{d\phi}{dt} = V_0 \cos\phi;$$

or eliminating t if $V_0 \neq 0$,

$$\frac{dr}{d\phi} = r\tan\phi + \frac{rV_r}{V_0\cos\phi}.$$

Substituting from Eq. (1) and rearranging, one gets for the relative path

$$r^4 \cos\phi \frac{dr}{d\phi} - r^5 \sin\phi - \frac{C_1}{V_0}r^3 - \frac{C_2}{V_0} = 0. \tag{3}$$

This equation is easily solved if either C_1 or C_2 is zero. Neglecting the dipole attraction, i.e., putting $C_2 = 0$, the solution of (3) is

$$r^2\cos^2\phi - \frac{2C_1}{V_0}\sin\phi = C.$$

For collision, $r = a_1 + a_2 = a$. This yields for the equation of the most distant path that leads to contact:

$$r^2\cos^2\phi = \frac{2C_1}{V_0}\sin\phi + a^2 + \frac{C_1^2}{V_0^2 a^2}$$

$$\text{if} \quad |C_1| \leqq |V_0| a^2. \tag{4.1}$$

$$r^2\cos^2\phi = \frac{2C_1}{V_0}\sin\phi - \frac{2C_1}{|V_0|} \quad \text{if } |C_1| \geqq |V_0| a^2. \tag{4.2}$$

Letting $\phi \to +\pi/2$ if $V_0 < 0$, and $\phi \to -\pi/2$ if $V_0 > 0$, we obtain for the effective radius of the charged particle,

$$a' = \lim r\cos\phi = a - \frac{C_1}{|V_0| a}, \quad \text{if } |C_1| \leqq |V_0| a^2. \tag{5.1}$$

$$a' = \lim r\cos\phi = \left(\frac{-4C_1}{|V_0|}\right)^{\frac{1}{2}}, \text{ if } |C_1| \geqq |V_0| a^2. \tag{5.2}$$

The volume swept in a time dt is given by

$$dV_{12}=\pi a'^2|V_0|dt=\pi\left(a^2|V_0|-2C_1+\frac{C_1^2}{a^2|V_0|}\right)dt=f_{12}'dt, \quad (6.1)$$

$$dV_{12}=\pi a'^2|V_0|dt=-4\pi C_1 dt=f_{12}''dt, \text{ respectively.} \quad (6.2)$$

Expression (6.2) is found to hold even for the extreme case of $V_0=0$.

If we do not neglect C_2 but put $C_1=0$, i.e., consider a charged particle encountering a neutral one, we obtain slightly more complicated expressions of a similar type. It is seen, however, that the volume swept in this case is small compared to the one considered first as soon as one charge is less than a hundred times larger than the other. We shall, therefore, in general be able to neglect C_2 altogether without introducing an appreciable error.

If the number of particles per unit volume that have a radius between a_2 and a_2+da_2 and a charge between q_2 and q_2+dq_2 is denoted by $n(a_2,q_2)da_2dq_2$ the number of such particles swept up by particle 1 in a time dt is given by

$$dv_{12}=n_{22}da_2dq_2dV_{12}=n_{22}f_{12}da_2dq_2dt,$$

abbreviating $n(a_2,q_2)$ by n_{22}.

The resulting increase in volume of particle 1 caused by particles 2 is then

$$d\Omega_{12}=4/3\pi a_2^3 n_{22}f_{12}da_2dq_2dt.$$

Integrating over all a_2 and q_2 that are encountered and expressing the growth with respect to radius, we get

$$\frac{da_1}{dt}=\frac{1}{3a_1^2}\int\int a_2^3 n_{22}f_{12}da_2dq_2. \quad (7)$$

Since n is largely unknown, this expression cannot be generally integrated. Making a few reasonable assumptions, however, one arrives at the conclusion that the rate of growth of a particle is negligible if

$$|qq_0|N/a_0<10^{10}, \quad (8)$$

where q is its charge in electrons, q_0 the average charge of the particles of one sign in electrons, a_0 the average radius of the particles in microns, and N the number of particles per cm³.

If that value is exceeded, aggregation might result in an increase of a_0 and a simultaneous decrease of q, q_0, and N until (8) is satisfied. The time taken for such a change will be of the order of a few seconds only.

The calculations involved are rather tedious and complicated, and will not be given here.

Since in most dusting processes q_0 easily exceeds 100, one cannot hope to maintain an N of the order of 10^6; in fact one should avoid such densities if one wishes to keep a_0 as small as possible.

Calendar of Meetings

November

15–17 American Oil Chemists Society, New York, New York
15–17 National Academy of Sciences, Berkeley, California
19–20 X-Ray and Electron Diffraction Conference, Detroit, Michigan
26–27 American Physical Society, Chicago, Illinois
28–December 3 American Society of Mechanical Engineers (Annual Meeting), New York, New York
29–December 1 Conference on Electronic Instrumentation in Nucleonics and Medicine (sponsored by A.I.E.E. and I.R.E.), New York, New York

December

5–8 American Society of Refrigerating Engineers, Washington, D. C.
7 Society of Applied Spectroscopy, New York, New York
12–13 American Chemical Society (Southwestern Regional Meeting), Houston, Texas
26–28 American Physical Society (Pacific Coast Meeting), Berkeley, California
26–31 American Association for the Advancement of Science, New York, New York
26–31 Mathematical Association of America, Columbus, Ohio
29–30 American Chemical Society, Division of Industrial and Engineering Chemistry, Chicago, Illinois

January, 1949

4 Society of Applied Spectroscopy, New York, New York
10–14 American Society of Mechanical Engineers, Conference on Materials Handling and Materials Handling Show, Philadelphia, Pennsylvania
27–29 American Physical Society, New York, New York

Magnitude and Character of Errors Produced by Shape Factors in Stokes' Law Estimates of Particle Radius

WULF B. KUNKEL*
Department of Physics, University of California, Berkeley, California
(Received May 20, 1948)

In order to estimate the possible errors introduced by deviations from the spherical particle shape in Stokes' law estimates of dust particle sizes, a model study was made with objects of various shapes falling in oil of high viscosity. It was found that all shapes fall more slowly than the sphere of the same mass and volume. The true size and mass will, therefore, always be larger than the estimates calculated from their rate of fall by means of Stokes' law. For shapes of *extreme* dimensions (very thin plates and needles) the error thus introduced exceeded 50 percent. For more common configurations of spheres it was of the order of 20 percent or less. No simple relation between particle surface or any other parameter and the rate of fall could be discovered. All plates and needles turned their planes and longer axes into a horizontal direction when falling freely, unless one side was distinctly weighted. A small asymmetry in weight distribution tilts the planes slightly, causing the model to drift sideways in its fall. In a Hopper and Laby method of particle analysis such a drift makes a particle appear charged, although it may be neutral. It is, therefore, not possible to analyze platelike and needlelike powder and dust particles by such a method.

IN the study of electrostatic properties of dusts and powders by means of the Hopper and Laby method for determining the charge on small particles, use is made of Stokes' law for spheres in the viscous medium. Since the shape of dust particles, however, will in general not be spherical, it was found necessary to investigate the probable deviations from Stokes' law caused by various shapes. As similar methods of investigation will undoubtedly be frequently applied, the following results were considered of a rather general interest and will, therefore, be briefly given here.

A glass cylinder 90 cm long and 10 cm wide was filled with a heavy white oil of known viscosity and density. The time taken by small objects of various shapes and sizes to fall between two accurately drawn marks was measured.

It was found that Stokes' law was accurately obeyed by small spheres of about 1 mm radius for velocities up to 8 cm/sec. The effect of the walls was not noticeable at distances larger than one cm from the walls. All this is in good agreement with the findings of other investigations.

To study the behavior of several spherical dust particles sticking together, glass beads were glued together to form small chains of two, three, etc., up to eight beads. Also plane arrangements of three and seven were tried and finally a small

TABLE I.

	No. of beads	Arrangement	m (mg)	$\bar{a}$ (mm)	t (sec.)	T (°C)	η (poises)	a_s (mm)	Difference in percent
linear	1	○	19.1	1.18	25.4	21.5	1.88	1.20	2
	2	○○	39.4	1.51	18.8	21.5	1.88	1.40	7
	3	○○○	58.5	1.72	17.1	20.9	1.98	1.50	13
	4	○○○○	85.1	1.95	15.6	21.0	1.95	1.56	20
	8	○○○○○○○○	161.0	2.40	13.9	21.2	1.92	1.64	32
plane	3		58.9	1.72	15.3	21.9	1.84	1.53	11
	7		137.9	2.28	10.7	21.9	1.84	1.83	20
space	6		110.7	2.12	10.4	22.0	1.83	1.85	13

* Acknowledgment is made to the Office of Naval Research, Contract No. N70NR295, T.O. 9, for assistance in this work.

cluster of six beads sticking together was dropped into the oil.

It was invariably found that groups fell more slowly than spheres of the same volume would, apparently because their surface was much larger. This means that a Stokes' law estimate of their size would always underestimate the real value. All arrangements of some symmetry about their major axis of *inertia* tended to line the latter up with the velocity of fall, thus offering the greatest possible resistance. Small straight chains would always turn their long axis into a horizontal plane and plate-like arrangements would always fall with their flat surfaces in a horizontal position. A slight asymmetry would tilt the major axis of inertia slightly, causing the particle to drift sideways in its fall. In a Hopper and Laby method of Stokes' law analysis of small charged particles, this deviation from vertical fall would make a particle appear charged while it is neutral. This can easily be checked by observing dust particles fall in the absence of a field. Such tests have revealed that some 10 percent of the particles in a fine quartz dust (1 to 10 microns diameter), work which was reported on at the January 2 meetings of the American Physical Society at Los Angeles, actually did not fall vertically.

Some typical values for the apparent size of little groupings are given in Table I. If m is the mass of the particle and ρ the density, the size is denoted by $\bar{a}$, the radius of the sphere of equal volume:

$$\bar{a}=\left(\frac{3m}{4\pi\rho}\right)^{\frac{1}{3}}=Am^{\frac{1}{3}}.$$

From the velocity we find a_s, the radius of a sphere that would obey Stokes' law as

$$a_s=\left(\frac{9}{2}\frac{\eta v}{(\rho-\rho_2)g}\right)^{\frac{1}{2}}=B\left(\frac{\eta}{t}\right)^{\frac{1}{2}}.$$

The glass beads were only approximately spherical; the viscosity depended on the temperature (T) and the velocity is inversely proportional to the time taken (t).

To find out whether there was any shape that would fall faster than the sphere of equal weight and thus appear in dust studies as being larger than it really was, pyrex beads were drawn out into points so that they formed tear-shaped models. These fell with their heavy ends pointing downward, but the velocity was again always slightly smaller than Stokes' law would predict (1 to 5 percent).

TABLE II.

Width (mm)	Weight (mg)	$\bar{a}$ (mm)	Surface (cm^2)	Time sec.	a_s mm	Difference in percent
5	215	1.85	7.3	14.4_5	0.81	56
			7.0	14.0_5	0.82	56
			6.6	13.5_0	0.83	55
			6.2	12.9_5	0.85	54
			5.8	12.4_0	0.87	53
			5.4	11.8_0	0.89	52
			4.7	10.9_5	0.93	50
			4.3	10.2_0	0.96	48
10	203	1.82	6.9	12.3	0.88	52
			6.5	11.9	0.89	51
			5.5	11.0	0.93	49
			4.0	9.4_5	1.00	45

It is thus seen that the deviation from Stokes' law increases with the deviation from the spherical shape and is always in the sense of causing slower fall than that of the sphere of same mass so that such size estimates would always yield a radius which is too small. Apparently the extent of the surface is of greater importance than the form of the streamlines.

Although the object of this survey was merely to get an impression of the magnitude and sign of the deviations to be expected from Stokes' law through shape factors, the measurements were carried a little further with special shapes of great simplicity.

Two long strips were cut out of thin steel foil so that their areas and weights were approximately equal, but their widths were in the ratio of one to two. One end was rolled in and squeezed flat to make the strips fall with that end pointing downward. By gradually rolling in more and more of the plate, the area could be varied without changing the weight. Since the foil was very thin (0.06 mm) compared to the width (5 mm and 10 mm), the total surface was approximately proportional to the length. In the region studied the time taken was found to be a linear function of the surface, the constants were, however, different for the two strips. Something of that nature was to be expected since the edge effect cannot be neglected in viscous flow no matter how sharp the edge is. No simple relation between

TABLE III.

Shape	m (mg)	Surface (one side) (mm²)	L (mm)	t (sec.)	η (poises)	δ (mm)	a_s (mm)	Difference in percent
○	7.1	15.1	13.7	77.0*	1.73	0.594	0.338*	43*
□	7.1	15.4	15.7	76.9			0.338	43
▭	7.1	15.4	18.4	83.9			0.326	45
○	28.1	60.8	27.6	42.0		0.940	0.458	51
□	28.1	60.8	31.2	42.5			0.456	51
▭	27.9	60.5	36.4	46.0			0.438	53
○	40.3	85.0	32.7	36.5		1.057	0.491	54
□	40.0	85.3	37.0	37.7			0.483	54
▭	40.0	85.3	42.8	40.1			0.469	56

* This value is not reliable since this smallest plate always drifted sideways approaching the wall of the container where it was probably slowed down. It apparently was not symmetrical enough, although no gross irregularity could be detected. This again shows how easily the direction of fall of a platelike particle will deviate from the vertical.

the length of the edge and the velocity could be discovered. Sample data are given in Table II. This yields the relation between time t and surface S,

$$t_5 = 1.43S + 4.1 \quad \text{for the 5 mm strip,}$$

and

$$t_{10} = 0.99S + 5.5 \quad \text{for the 10 mm strip.}$$

It is seen that the greatest difference between Stokes' law radius and actual radius of equivalent sphere in these cases was 56 percent.

To demonstrate the effect of the shape at constant surface, small plates were cut of the same material as before, but this time circular, square, and rectangular shapes with equal surfaces (and of course equal weights) were chosen. Three of these sets were made in different sizes. No simple relation between rate of fall and total edge could be discovered, but the time taken increased with the length of the total edge as one could have expected. The data are given in Table III. The temperature was constant in this experiment. The total length of edge is denoted by L.

Again the largest difference observed between Stokes' law and true size was about 56 percent.

One may, in general, conclude that the shape and surface of particles influences the rate of all under Stokes' law in such a sense as always to give a smaller velocity than that of a sphere of equivalent weight. This means that estimates of particle size from the rate of fall will always be in such a sense as to give radii and apparent masses that are smaller than those of the equivalent sphere. The maximum error introduced will depend on the shapes involved. Since in most powders and dusts such extreme shapes as very thin plates or needles are rather improbable, it is fairly safe to say that the maximum error will, in general, be less than 50 percent. It is, however, well advised to study the shape of the particles before they are subjected to an analysis by means of Stokes' law.

On the Measurement of Cavity Impedance

W. W. Hansen and R. F. Post
Stanford University, Stanford, California
(Received May 17, 1948)

A method of measuring cavity impedance is described.

INTRODUCTION

To describe the properties of a resonant cavity, operating in a specified mode, at least three constants are needed: two that are almost always useful are λ and Q. The third may be the shunt resistance R or the quantity R/Q. The former is usually of most direct interest; the latter is sometimes convenient because it depends only on resonator geometry and not on material or loading.

For present purposes, R/Q is the more convenient and may be defined as

$$R/Q=2c\frac{\lambda\left[\int E\cdot ds\right]^2}{\int B^2 d\tau}=2c\frac{E^2s^2\lambda}{\int B^2 d\tau}\left[\frac{\int E\cdot ds}{E\cdot s}\right]^2, \quad (1)$$

where E and B are the peak fields and the integral of E is over the path s along which one wishes to measure shunt impedance. The third member of the equality is identical with the second, and the E occurring outside the integral means E at some specific point on s, for example, the cavity wall. This form is convenient if one can find a way of measuring this E in terms of $\int B^2 d\tau$ and can also devise means for measuring the ratio $\int E\cdot ds/E\cdot s$.

Of course, the definition (1) is, to a considerable extent, arbitrary and it is often desirable to modify it, usually by introducing various weight factors in the integral of $E\cdot ds$. Ordinarily these modifications will not affect the utility of what follows.

To see how to measure R/Q, consider the cavity of Fig. 1a, which might be suitable for use in a klystron.[1] Suppose we measure the change in resonant wave-length, call it $\Delta\lambda$, caused by the introduction of conducting "buttons" of arbitrary shape and volume $\Delta\tau$, for example, the cylinder of radius r and height d shown in dotted lines. Then, by variation methods, or otherwise, one can show that $\Delta\lambda$ is given "approximately" by

$$\frac{\Delta\lambda}{\lambda_0}=\frac{1}{2}\frac{\int_{\Delta\tau}(E^2-B^2)d\tau}{\int B^2 d\tau}, \quad (2)$$

where λ_0 is the unperturbed wave-length, E is the unperturbed field at the button, $\Delta\tau$ is the volume of this button, and the integral of B^2 in the denominator is over the entire cavity.

Now, one might think of measuring $\Delta\lambda$ and $\Delta\tau$ for buttons sufficiently small to make the approximations good and so determine the ratio $(E^2-B^2)/\int B^2 d\tau$. Then, for small buttons on the axis, the B^2 in the bracket is negligible, or may be corrected for, and so the right-hand side of Eq. (1) may be evaluated, the ratio $\int E\cdot ds/Es$ being not too hard to find.

This is, in fact, the procedure here adopted, but to make it work, several observations must be made.

First, Eq. (2) is correct, even in the limit of vanishing $\Delta\tau$, only for those shapes of perturbing volume or button which do not significantly alter the field at large distances from the button. To say the same thing in another way, the value approached by $\Delta\lambda/\Delta\tau$ depends on the manner

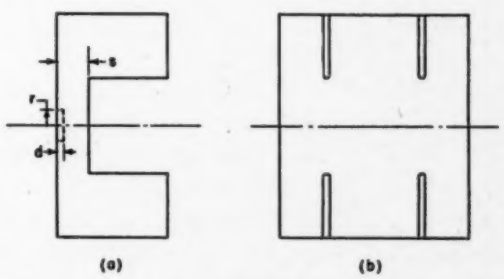


Fig. 1. (a) Reëntrant cylindrical cavity. (b) Disk-loaded cylindrical cavity.

[1] R. L. Sproull and E. G. Linder, Proc. IRE **34**, 305 (1946) give another method of measurement which is applicable when E is constant.

in which $\Delta\tau$ approaches zero. For example, if cylindrical buttons are used and $d\to 0$ first, then Eq. (2) is correct, but if $r\to 0$ first, the right-hand side must be multiplied by an infinite factor. Again, if one uses a half-sphere of radius r the right-hand side must be multiplied by 3. To obtain correct results, then, one must either use a button shape for which Eq. (2) is correct or must find the appropriate factor required to make the formula correct. This might be done theoretically or by trial on a calculable cavity.

If this be done, correct results are possible in principle, but there may be serious practical troubles because of inability to find a compromise button size which is small enough not to perturb the cavity fields, and at the same time large enough to give accurately measurable wave-length shifts. For example, consider the cavity of Fig. 1b. The shapes of interest to us are $\lambda/2$ long and about λ in diameter to give a volume $\cong\lambda^3/2$. Then, if we aim at a fractional error ϵ and use a hemispherical button we must have, approximately, $8r/\lambda<\epsilon$, since the distance from button to loading disk is $\cong\lambda/8$. But this makes $\Delta\lambda/\lambda \sim\epsilon^3/2$ and with $\epsilon=10^{-2}$ this would require considerable accuracy in wave-length measurement.

Of these difficulties, the first may be avoided entirely, and the second may be greatly reduced, by the almost obvious procedure of using a cylindrical button and taking the $\lim d\to 0$ first. And this limiting procedure is greatly increased in accuracy by taking data for both positive and negative d. Then one plots $\Delta\lambda$ *vs.* d and takes the slope at $d=0$, so using Eq. (2) for a case where it surely applies, since the fields are completely unperturbed.

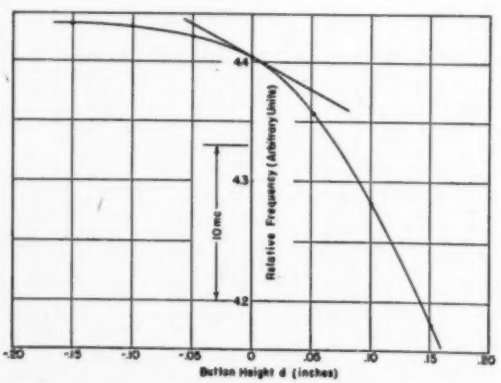


FIG. 2. Resonant frequency *vs* button height—cylindrical cavity.

In this way one can determine $\int(E^2-B^2)d\sigma/\int B^2 d\tau$ where the surface integral, which is over the face of the button, is nearly equal to $\pi r^2 E_0^2$ with E_0 the field on the axis. The equality could be approached as nearly as one wishes by reducing r, were it not for the resulting reduced accuracy. In many cases, one may use buttons of moderate size by making use of the known properties of the field to evaluate a suitable correction factor. Thus, in the present case where we have a field with axial symmetry, bounded by a plane surface, B_φ may be well approximated by $J_1(k_1\rho)\cos k_3 z$, and E_z by the appropriate related fields, and so we can find, for this case

$$\int(E^2-B^2)d\tau\cong\pi r^2 d[1-r^2(\tfrac{3}{8}k^2-\tfrac{1}{4}k_3^2)]E_0^2, \quad (3)$$

where E_0 is the field on the axis. The quantity k_3 is determined in the course of the measurements of the axial dependence of the field.

Finally, combining (1), (2), and (3), we have

$$R/Q=2c\frac{s^2}{\pi r^2}\frac{\delta\lambda}{\delta d}\frac{1}{[1-(\frac{3}{8}k^2-\frac{1}{4}k_3^2)r^2]}\times\left[\frac{\int E\cdot ds}{E\cdot s}\right]^2, \quad (4)$$

where $\delta\lambda/\delta d$ is the slope of the λ *vs.* d curve at $d=0$.

To evaluate the last factor one needs knowledge of the variation of E along the curve s. This is easily obtained[2] by moving a bead of dielectric along the curve s and plotting the wave-length shift *vs.* position of the bead. In some cases, for example Fig. 1a, it is obvious from the geometry that E is nearly constant. In these circumstances this factor becomes unity.

[2] This method is mentioned on page 59 of a "Quarterly Progress Report" of the Research Laboratory of Electronics, Massachusetts Institute of Technology, dated January 15, 1947. Another method might appear to be to observe the wave-length shift produced by a thin dielectric rod along the path s. This would seem to measure $\bar{E}^3$ which would be nearly equal to $\bar{E}^2$. Unfortunately when E is not constant the dielectric changes the field distribution by an amount that is generally not negligible, for reasons similar to those discussed in connection with Eq. (2). This point was also noted by Sproull and Linder.

EXPERIMENTAL

We describe in this section some of our experiences with the procedure sketched above.

As a check on the absolute accuracy of the method, measurements were made on two computable cavities of cylindrical form. One of these had a diameter of $3\frac{1}{4}''$ and a length of $3\frac{1}{2}''$ and resonated in the mode with no axial nodes at at 2770 mc. The other had a diameter of $3\frac{3}{4}''$ and length of $3''$ and resonated near 2990 mc in the mode with one node in the axial direction. In each case one of the plane ends was removable and could be replaced by one of a sequence of plates with "buttons" of heights $d=\pm 0.050''$, $\pm 0.100''$, $\pm 0.150''$, and all of diameter $2r=0.750''$. A typical curve of frequency *vs.* d is shown in Fig. 2 and from such we determined R/Q. Similar experiments were done with $2r=0.500''$. In all cases the results agreed with the theory, to within 4 percent or better. We regard this as satisfactory proof that the method is sound.

The only experimental difficulty encountered with this method was in obtaining sufficient reproducibility in the shape (flatness) of the changeable end plates. This is easy to understand since a $\frac{3}{4}''$-diameter button has $\sim\frac{1}{16}$ the area of the $3\frac{1}{4}''$-end plate so that if, say, five thousandths on the button height are to be significant, the plate must be flat to about $\frac{1}{10}$ of this, or $0.0005''$. And this tolerance includes any distortions caused by clamping. Once this elementary source of error was understood, no great difficulty was found in eliminating it and obtaining the data reproduced above.

Another difficulty is that of determining the slope of a graph as highly curved as that of Fig. 2.[3] This trouble is more apparent than real, experience showing that derivatives as obtained graphically, by a curve-fitting formula, and by extrapolating the slopes of chords drawn between points symmetric about $x=0$, have substantially the same value.

If it is desired to avoid the slight inconvenience or uncertainty caused by this curvature, one has only to observe that the wave-length shift $\Delta\lambda$ may be written approximately as $\Delta\lambda=\alpha f(d/r)$, where α depends on the ratio of the field at the button to the energy stored and f is a universal function. Thus, if we measure $\Delta\lambda(d/r)$ for some chosen cavity, and call it $\Delta_0\lambda$, then a plot of $\Delta\lambda$ for some other cavity against $\Delta_0\lambda$ will give a straight line with slope determined by the ratio of $E^2\lambda^3/\int B^2 d\tau$ for the two cavities. Such a plot is shown in Fig. 3, the data being for two widely different cavities of the general form of Fig. 1b. These data were taken with a micrometer-controlled button to be described in the next paragraph, and the points correspond to values of d differing by $0.012''$.

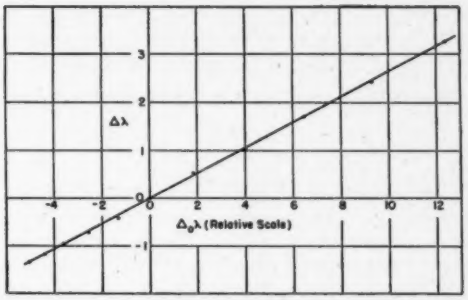


FIG. 3. $\Delta\lambda$ *vs.* $\Delta_0\lambda$. Cavities of type shown in Fig. 1b.

Encouraged by the success of the above-described preliminary experiments, we made an end plate with a $0.750''$ "button" moved by a micrometer screw and with connection between the "button" and the plane surface established by a carefully made choke joint of low impedance. This apparatus speeds the work greatly, increases the accuracy of measurement of d, and eliminates errors resulting from flexure and imperfect flatness of the plates.

This worked very well on cavities of the form shown in Fig. 1b, the data looking much better than that of Fig. 2. We have also made a similar but smaller movable plunger and have used it to measure a number of cavities of the type of Fig. 1a and with proportions useful in klystrons. Judging the results by trial on exactly calculable cases, by deviations from smooth curves, or by comparison with numbers obtained by careful net point calculations, it appears that answers good to ± 5 percent may be obtained without difficulty.

As a result of the above-described experiments, we have concluded that this method of measurement of characteristic impedance is practical, rapid, and reasonably accurate.

We wish to acknowledge with thanks the support of the Office of Naval Research.

[3] It is apparent from the curvature of Fig. 2 that the doubts expressed about the validity of Eq. (2) for finite perturbations are amply justified.

The Fracture of Liquids

JOHN C. FISHER
General Electric Research Laboratory, Schenectady, New York
(Received June 21, 1948)

A liquid subjected to negative pressure is metastable; vapor bubbles form spontaneously and grow until the pressure of the system rises to the equilibrium vapor pressure. The rate of bubble formation is calculated from the theory of nucleation, and the negative pressure p_t that gives one bubble (i.e., fracture) in t seconds is determined. p_t is very nearly independent of t, and is proportional to σ^3 where σ is the surface tension. Subcooled liquids such as glass also are metastable under negative pressure; cracks form spontaneously and grow until the pressure rises to the equilibrium vapor pressure. Nucleation theory leads to an expression for the fracture stress of glass that is proportional to $(E^2\sigma^3)^{\frac{1}{5}}$ where E is the elastic modulus. The transition from high temperature cavity-nucleated fracture to low temperature crack-nucleated fracture is examined.

Fracture strengths calculated from nucleation theory agree satisfactorily with the maximum experimental values, and are an order of magnitude smaller than the forces required for simultaneous separation of all atomic bonds cut by a plane surface. The frequent occurrence of premature failure is associated with the presence of pre-existing surface cracks in glass, and of positive contact-angle impurity in contact with liquids.

THE FRACTURE OF LIQUIDS

LIQUIDS frequently are observed to withstand negative pressures (hydrostatic tensile stresses) of considerable magnitude. Yet a liquid subjected to negative pressure is metastable; in time it will change spontaneously to the two phase system liquid plus vapor, and the pressure will rise to the equilibrium vapor pressure. Cavitation, or the spontaneous formation of vapor in a stressed liquid, is of interest to engineers who wish to avoid it in the operation of underwater sound generators, ship's propellers, and other underwater equipment. A satisfactory analysis of cavitation problems requires a reliable estimate of the maximum negative pressure that a given liquid can withstand, as well as an examination of the conditions that may reduce this value in practice.

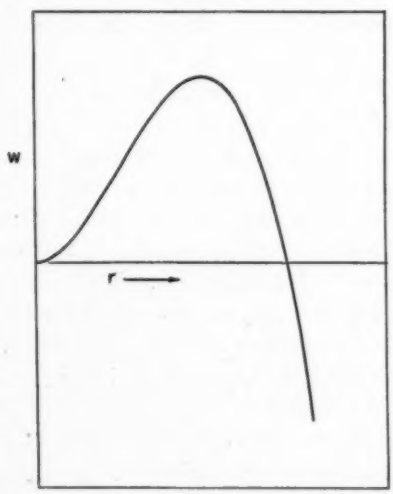


FIG. 1. Work required for reversible formation of a bubble of radius r in a liquid under negative pressure.

The force required to separate a specimen of material into two portions along a plane surface has frequently been taken as the theoretical fracture strength value. It is assumed in such calculations that all parts of both plane faces of the fractured specimen appear at the same time. The vapor bubbles and cracks responsible for the failure of liquids and amorphous solids under tension are observed on the contrary to grow from extremely small nuclei. Application of the methods of nucleation theory to the nucleation of vapor bubbles in stressed liquids[1,2] and of cracks in stressed solids leads, therefore, to more realistic estimates of the theoretical fracture strengths of materials.

A definite quantity of work is associated with the reversible formation of a vapor bubble in the interior of a liquid. Formation of a cavity of volume V requires work equal to pV where p is the pressure of the liquid. When the liquid is under negative pressure, the work pV is negative. Formation of the liquid-vapor interface bounding the bubble requires work equal to σA, where σ is the liquid-vapor interfacial tension and A is the area of the interface. The work required to fill the

[1] M. Volmer, *Kinetik der Phasenbildung* (Theodor Steinkopff, Dresden and Leipzig, 1939).
[2] J. Frenkel, *Kinetic Theory of Liquids* (The Clarendon Press, Oxford, 1946).

bubble reversibly with vapor of pressure p_r is negative and equal to $-p_rV$. The net work associated with the reversible formation of a spherical vapor bubble of radius r is, therefore,

$$W=\sigma A+pV-p_rV$$
$$=4\pi r^2\sigma+(4/3)\pi r^3(p-p_r). \quad (1)$$

The present analysis treats the nucleation of vapor bubbles in liquids under high hydrostatic tension. For negative pressures of this magnitude, p_r is negligible in comparison with p, and the work expression simplifies to

$$W=4\pi r^2\sigma+(4/3)\pi r^3p. \quad (2)$$

W is plotted *versus* r in Fig. 1 for a negative value of p. The curve has a maximum $W_{max}=16\pi\sigma^3/3p^2$ for bubbles with radius $r^*=-2\sigma/p$. Bubbles with radii less than r^* require free energy for further growth, while those with radii larger than r^* grow freely with decreasing free energy. Since bubbles grow one atom at a time as the result of statistical thermal fluctuations, it is evident that small bubbles with radii less than r^* will usually disappear without reaching the critical size r^*. Only exceptionally will a long chain of favorable energy fluctuations produce a bubble exceeding the critical size r^*. When this rare event does happen, however, the supercritical bubble will grow until the pressure of the system rises to the vapor pressure of the liquid.

The theory of nucleation[1-3] states that the rate of bubble formation is proportional to $\exp(-W_{max}/kT)$. The proportionality factor can be estimated from the theory of absolute reaction rates to be $(nkT/h)\exp(-\Delta f_0^*/kT)$, where n is the number of molecules in the liquid, and Δf_0^* is the free energy of activation for the motion of an individual molecule of liquid past its neighbors into or away from the bubble surface.[4] The rate of formation of bubbles of vapor in a mole of liquid subjected to negative pressure p is, therefore,

$$\frac{dn}{dt}=\frac{NkT}{h}\exp[-(\Delta f_0^*+W_{max})/kT]$$
$$=\frac{NkT}{h}\exp[-(\Delta f_0^*+16\pi\sigma^3/3p^2)/kT] \quad (3)$$

[3] Fisher, Hollomon, and Turnbull, J. App. Phys. 19, 775 (1948).
[4] D. Turnbull and J. C. Fisher, to be published.

TABLE I. Calculated fracture pressure p_t for water at 300°K. (10^9 dynes/cm^2)

Waiting time t (sec)	Free energy of activation (Kcal./mole) $\Delta f_0^*=0$	$\Delta f_0^*=5$	$\Delta f_0^*=10$
10^{-15}	−1.74	−1.91	−2.14
10^{-12}	−1.63	−1.77	−1.94
10^{-9}	−1.54	−1.65	−1.80
10^{-6}	−1.46	−1.56	−1.68
10^{-3}	−1.40	−1.48	−1.58
1	−1.34	−1.41	−1.50
10^3	−1.29	−1.35	−1.43
10^6	−1.24	−1.30	−1.36
10^9	−1.20	−1.25	−1.31
10^{12}	−1.16	−1.21	−1.26
10^{15}	−1.13	−1.17	−1.22
10^{18}	−1.10	−1.13	−1.18

(assuming, of course, that somehow the pressure is kept at p even after the first bubbles have begun to grow).

It is fairly easy to estimate from Eq. (3) the negative pressure at which a liquid will fail. Since the first bubble that forms fractures the liquid, the fracture pressure will be the negative pressure that gives one bubble in a reasonable time—say one bubble in t seconds. Then $dn/dt=1/t$ and Eq. (3) can be solved for the fracture pressure p_t,

$$p_t=-\left[\frac{16\pi}{3}\frac{\sigma^3}{kT\ln NkTt/h-\Delta f_0^*}\right]^{\frac{1}{2}}. \quad (4)$$

The subscript t is added to the symbol for the fracture pressure p_t to emphasize that the value of p_t depends upon the time that one is willing to wait for the first vapor bubble to appear.

Values of the fracture pressure p_t corresponding to a number of waiting times and to three values of Δf_0^* are given in Table I for water at 300°K. All time values of possible interest are covered, from $t=10^{-15}$ sec. (less than the time required for sound to travel one atomic distance) to $t=10^{18}$ sec. (exceeding the estimated age of the universe). When $\Delta f_0^*=0$, for example, the ratio of the maximum to the minimum fracture pressure is only 1.58, although the corresponding time ratio is 10^{33}. This extreme insensibility of the fracture pressure value suggests that the fracture pressure p_1 corresponding to one vapor bubble per mole per second cannot be appreciably in error for tests of practical duration.

The value of the free energy of activation Δf_0^* for the motion of a molecule into or away from the bubble surface can be estimated from the free

TABLE II. Fracture pressure p_1 for several liquids at approximately 300°K.

Liquid	σ (ergs/cm²)	p_1 (atmospheres) Calculated	Maximum experimental
Mercury	474	−22,300	—
Water	72	−1320	−350[a] −120[b] −42[c]
Ethyl alcohol	22	−223	−39[d]
Ethyl ether	16	−138	−72[d]

[a] A. Ursprung, Ber. D. Bot. Ges. **33**, 153 (1915).
[b] 5 mc ultrasonic radiation. Fracture not obtained. J. F. Muller and G. W. Willard, presented at 35th meeting of Acous. Soc. Am., April 22–24, 1948.
[c] H. N. V. Temperley, Proc. Phys. Soc. London **58**, 436 (1946).
[d] Later work indicates the possibility of calibration errors. J. Meyer, *Zur Kenntniss des negativen Druckes in Flüssigkeilen* (W. Knapp, Halle, 1911).

energy of activation Δf_{vis}^* for viscosity, as the two free energies should be approximately equal. Values of Δf_{vis}^* calculated by Roseveare, Powell and Eyring[5] lie between 0 and 5000 cal./mole for most liquids that are fluid at room temperature, and are less than 7500 cal./mole for all such liquids investigated. The relative variation of p_t with Δf_0^* is independent of the liquid under consideration, and an examination of Table I indicates that putting $\Delta f_0^* = 0$ will give a value of p_t generally less than 5 percent and almost certainly less than 10 percent too small for room temperature liquids.

Nucleation theory leads therefore to a rather definite fracture pressure for fluid liquids,

$$p_1 = -\left[\frac{16\pi}{3kT}\frac{\sigma^3}{\ln NkT/h}\right]^{\frac{1}{3}}. \qquad (5)$$

Theoretical fracture pressure values calculated from Eq. (5) are listed in Table II for several liquids at 300°K. It is of interest to note that these values are an order of magnitude smaller than those predicted from the forces required to separate the material across a plane surface.

Glass and other undercooled liquids may fail by the nucleation and propagation of cracks, rather than of bubbles as do more mobile liquids,[6] reversible formation of the two surfaces of a circular crack requires work equal to $2\pi r^2\sigma$ where σ is the solid surface tension and r is the radius of the crack. The work required to allow the reversible relaxation of strain energy in the neighborhood of the crack has been shown by Sack to be $-8(1-\nu^2)r^3s^2/3E$ where s is the stress in the direction normal to the plane of the crack.[7] The net work associated with the reversible formation of a circular crack of radius r is, therefore,

$$W = 2\pi r^2\sigma - 8(1-\nu^2)r^3s^2/3E. \qquad (6)$$

The curve of W *versus* r has a maximum $W_{max} = \pi^3\sigma^3E^2/6(1-\nu^2)s^4$ for cracks with radius $r^* = \pi\sigma E/2(1-\nu^2)s^2$. Nucleation theory leads to

$$\frac{dn}{dt} = \frac{NkT}{h}\exp[-(\Delta f_1^* + W_{max})/kT]$$

$$= \frac{NkT}{h}\exp[-(\Delta f_1^* + \pi^3\sigma^3E^2/6(1-\nu^2)s^4)/kT] \qquad (7)$$

for the rate of nucleation of cracks in a mole of undercooled liquid subjected to a hydrostatic tensile stress s. Δf_1^* represents the free energy of activation for separating a pair of atoms as the

TABLE III. Fracture stress s_1 for several undercooled liquids.

Material	σ (ergs/cm²)	E (10^9 dynes/cm²)	s_1 (10^9 dynes/cm²) Calculated	Maximum experimental
Glass	~150–300[a]	400–800	30–72	34[b]
Pyrex glass	~250[a]	610	55	120[c]
Fused quartz	~350[a]	700	76	90[d]

[a] Estimated from the surface tension measurements of E. W. Washburn, G. R. Shelton and E. E. Libman, U. of Illinois Eng. Experiment Station Bull. No. 140 (1924).
[b] A. A. Griffith, Phil. Trans. Roy. Soc. Lond. **A211**, 163 (1921); Proceedings of the First International Congress for Applied Mechanics (Delft, 1924).
[c] S. Jurkov, J. Tech. Phys. USSR **1**, 386 (1935).
[d] O. Reinkober, Physik. Zeits. **32**, 243 (1931).

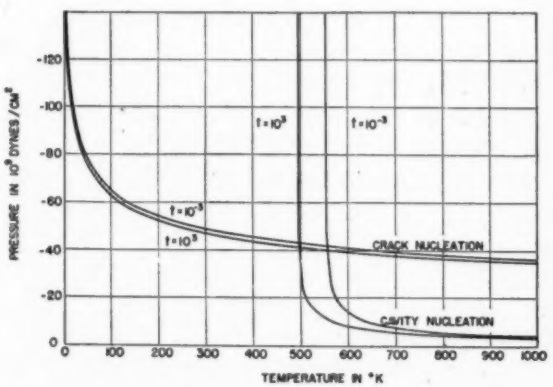


FIG. 2. Negative pressure required for nucleation of a bubble or crack in glass.

[5] Roseveare, Powell, and Eyring, J. App. Phys. **12**, 669 (1941).
[6] E. F. Poncelet, Metals Technology (1944).
[7] R. A. Sack, Proc. Phys. Soc. London **58**, 729 (1946).

edge of the crack moves between them. The fracture stress corresponding to the nucleation of one fracture crack per mole per second is

$$s_1=\left[\frac{\pi^3E^2}{6(1-\nu^2)}\frac{\sigma^3}{kT\ln NkT/h-\Delta f_1^*}\right]^{\frac{1}{3}}. \quad (8)$$

Although this expression has been derived for hydrostatic tension, it will be assumed to hold for simple tension and other combined stress states with one or more positive principal stresses. s_1 will be identified with the value of the maximum tensile stress corresponding to the nucleation of one crack per second in a mole of material.

The value of Δf_1^* can be estimated from the free energy of activation Δf_{dif}^* for self-diffusion, suitably modified to correspond to an applied stress equal to the fracture stress s_1. Obviously s_1 cannot be less than the value corresponding to $\Delta f_1^*=0$ in Eq. (8). Fracture stress values corresponding to this assumption are tabulated in Table III for glass and fused quartz. The minimum values of s_1 calculated assuming $\Delta f_1^*=0$ prove to be so large that the corresponding true values of Δf_1^*, which decrease with increasing applied stress, are assumed to be negligibly small.

Nucleation theory predicts, therefore, a rather definite hydrostatic tensile stress for fracture of non-crystalline solids by the nucleation and propagation of cracks,

$$s_1=\left[\frac{\pi^3E^2}{6(1-\nu^2)kT}\frac{\sigma^3}{\ln NkT/h}\right]^{\frac{1}{3}}. \quad (9)$$

For very fluid liquids there is little doubt that failure in hydrostatic tension will occur by the nucleation and growth of a spherical cavity or bubble. Similarly, for very greatly subcooled liquids such as fused quartz at room temperature there is little doubt that failure will occur by the nucleation and growth of a crack. However, for liquids of high viscosity such as glass, in the transition from liquid to solid behavior, failure may occur by either mechanism. In locating the transition from cavity-nucleated fracture to crack-nucleated fracture it is necessary to employ Eq. (4) for the rate of cavity nucleation. The approximating Eq. (5) is invalid for regions of high viscosity or low temperature where $\Delta f_0^*/kT$ is comparable in magnitude with $\ln NkTt/h$.

TABLE IV. Properties of a typical soda-lime-silica glass

$E=6.5(10)^{11}$ dynes/cm² [a]
$\nu=0.25$ [a]
$\sigma=200$ ergs/cm² [b]
$\Delta f_{vis}^*/kT=-35+62{,}500/T$ [c]

[a] International Critical Tables.
[b] See reference (a), Table III.
[c] Estimated from the variation of viscosity with temperature according to the method of reference 5, using the experimental data of reference (a), Table III.

Calculation of the temperature of transition from the cavity type failure to the crack type failure can be carried out most easily for glass, as this material has been studied extensively and its properties are known. Table IV gives representative values for certain properties of a typical soda-lime-silica experimental glass, taken from several sources.

The stress for failure by cavity nucleation calculated by means of Eq. (4) is plotted *versus* temperature in Fig. 2 for two fracture times. The stress for failure by crack nucleation calculated by means of Eq. (9) is also plotted *versus* temperature for the same fracture times. For each fracture time there is a transition from a low-temperature crack-nucleated fracture to a high-temperature cavity-nucleated fracture. The transition temperature T_e is determined very nearly by the relationship

$$\ln NkT_et_e/h=\Delta f_0^*/kT_e \quad (10)$$

where t_e is the time to fracture.

Comparison of the calculated and experimental fracture stress values given in Table II indicates that the fracture stresses calculated from nucleation theory generally exceed the maximum experimental stresses. Measured values of the fracture pressure of water, for example, range from the positive vapor pressure to a negative pressure of about 350 atmospheres. However, the theoretical fracture pressure of 1320 negative atmospheres was derived assuming that the vapor bubble responsible for fracture formed in the interior of the liquid. The low experimental values may result from the fact that the initial vapor bubble forms instead at the interface between the liquid and the container, or between the liquid and a suspended particle.

A vapor bubble at the interface between a liquid and a plane solid surface assumes a shape

TABLE V. Pressure for fracture at a plane liquid-solid interface.

Liquid—Solid	Contact angle	p_{i1} (atmospheres)
Water—Paraffin wax	105°[a]	−820
Water—Crude montan wax	140°[b]	−290
Water—½ Stearone+½ shellac wax	155°[b]	−120
Mercury—Glass	140°[a]	−4800
Mercury—Steel	154°[a]	−2100

[a] N. K. Adam, *The Physics and Chemistry of Surfaces* (Oxford University Press, London, 1941).
[b] Probably not a smooth surface. R. N. Wenzel, Ind. Eng. Chem. **28**, 988 (1936).

bounded by a plane and a portion of a spherical surface, as shown in Fig. 3. In the figure, σ_L, σ_S and σ_{SL} represent the liquid-vapor, solid-vapor and solid-liquid surface tensions, respectively. The value of d corresponding to minimum free energy can be shown to be

$$d/r=m=(\sigma_{SL}-\sigma_S)/\sigma_L. \tag{11}$$

The work required for reversible formation of the bubble illustrated in Fig. 3 can be calculated as the sum of three terms,

$$W=2\pi r^2(1-m)\sigma_L+\pi r^2(1-m^2)(\sigma_S-\sigma_{SL}) -(\pi r^3/3)(2-3m+m^3)p. \tag{12}$$

Here $2\pi r^2(1-m)$ is the area of the liquid-vapor interface, $\pi r^2(1-m^2)$ is the area of the solid-liquid interface, and $(\pi r^3/3)(2-3m+m^3)$ is the volume of the bubble.

The maximum value of W, corresponding to a bubble of critical size, is

$$W_{\max}=16\pi\sigma_L{}^3\phi/3p^2 \tag{13}$$

where

$$\phi=(2+m)(1-m)^2/4. \tag{14}$$

The corresponding pressure for fracture at the interface is

$$p_{i1}=-\left[\frac{16\pi}{3kT}\frac{\sigma_L{}^3\phi}{\ln 6N^{\frac{1}{3}}kT/h}\right]^{\frac{1}{3}}, \tag{15}$$

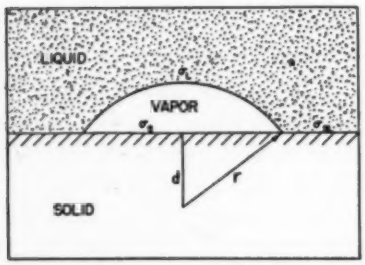


FIG. 3. Vapor bubble at a solid-liquid interface.

which differs from Eq. (5) in that a factor of $\phi^{\frac{1}{3}}$ has been introduced and the number of interface atoms $6N^{\frac{2}{3}}$ (assuming a cubical container) appears in place of N in the logarithm.

In Table V the negative pressure for fracture of water at a solid-liquid interface is compared with that for fracture in the interior of the liquid. The pressure p_{i1} for fracture at the interface is seen to depend upon the nature of the solid material. The interface fracture pressure of water in contact with a plane paraffin surface is, for example, only 62 percent of that for fracture in the interior of the liquid, even though the number of possible fracture sites has been reduced.

There are sound reasons for expecting the experimental fracture pressures of liquids in contact with solids to be very much less numerically than the calculated values given by Eq. (15). Consider, for example, water in contact with a paraffin surface having a conical cavity as sketched in Fig. 4. It is evident that no finite pressure can force the water to the bottom of the cavity; therefore, there is no problem of nucleating the vapor phase when the pressure is reduced. The small vapor region at the apex of the cavity will grow indefinitely as soon as the pressure drops below the vapor pressure of water. If the apex of the cavity is rounded, as sketched in Fig. 5, a finite pressure will fill the cavity. However, the nucleation of a vapor bubble at the rounded apex cannot require work in excess of

$$W'=(\sigma_S+\sigma_L-\sigma_{SL})A$$

where A is the apex area, and for sufficiently small A the fracture pressure corresponding to bubble nucleation at the apex of the cavity can approach the vapor pressure as closely as desired.

Actual solid surfaces are rough and irregular. If the liquid does not have zero contact angle with the solid, a wide range of experimental fracture pressure values should be observed. If the liquid has not been forced to the bottoms of all cavities in the solid by a prior application of pressure, vapor bubbles will appear at the vapor pressure of the liquid. If, on the other hand, the liquid has been forced to the bottoms of all cavities by prior pressure, a vapor bubble will form at the apex of the severest cavity when the pressure reaches a value between the vapor pressure and that corresponding to Eq. (15). If the liquid has zero con-

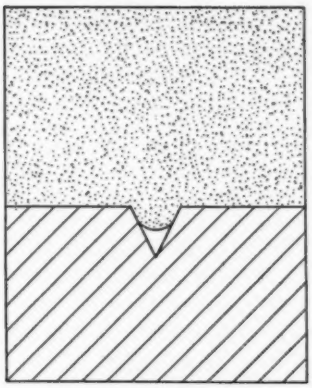


FIG. 4. Water in contact with a paraffin surface having a conical cavity.

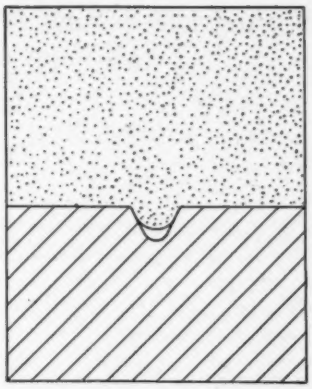


FIG. 5. Water in contact with a paraffin surface having a conical cavity with a rounded apex.

tact angle with the solid, as does water in contact with glass, it is sufficient to postulate the existence of small regions of positive contact-angle material or positive contact-angle impurity to obtain an equally wide range of experimental fracture pressures. The variation of observed fracture pressures can be interpreted satisfactorily in this manner.

The maximum observed fracture stress values for glass and fused quartz agree very well with the theoretical values calculated from nucleation theory. The low fracture strengths associated with glass specimens that have not received special protection against surface damage can be shown to result from accidental surface cracks.[8] Light contact between two glass rods produces cracks that are much larger than the critical size. Atmospheric attack causes such surface cracks to persist when they might otherwise have healed. Large surface cracks grow under small applied stresses, leading to the usual premature failure of glass. The marked time effects frequently observed in the fracture of glass are believed to depend upon the diffusion of surface films of water and other substances into surface cracks, and upon the associated changes in surface tension.[9]

It is interesting to note that glass specimens probably fail by surface cracks even in the absence of surface damage. Assuming that a semi-circular surface crack relieves half the strain energy relieved by an internal circular crack of the same radius, the fracture stress corresponding to the nucleation of one surface crack per second per mole of material is

$$s_{s1}=\left[\frac{\pi^3E^2}{12(1-\nu^2)kT}\frac{\sigma^3}{\ln 6N^1kT/h}\right]^{\frac{1}{2}}, \qquad (16)$$

a value that is a few percent less than s_1 as given by Eq. (9).

Nucleation theory has led to theoretical values for the fracture stresses of liquids and non-crystalline solids that are an order of magnitude smaller than the values calculated from the forces required for the simultaneous separation of all the atomic bonds cut by a plane surface. The maximum experimental fracture stress values approach those derived from nucleation theory, suggesting that the theoretical fracture strengths are now more nearly correct.

The frequent occurrence of premature fracture has been associated with the presence of surface irregularities. It is well known that fracture cracks in glass grow most easily from pre-existing surface cracks. In liquids, similarly, vapor bubbles are much more easily nucleated at an irregular solid-liquid interface than in the interior of a liquid, and the fracture strength corresponding to failure at an interface can have any value numerically less than the theoretical strength of the liquid.

[8] See reference (b), Table III.

[9] T. C. Baker and F. W. Preston, J. App. Phys. **17**, 179 (1946).

Geometrical Factors Affecting the Contours of X-Ray Spectrometer Maxima. I. Factors Causing Asymmetry*

LEROY ALEXANDER
Department of Research in Chemical Physics, Mellon Institute, Pittsburgh, Pennsylvania
(Received June 11, 1948)

The observed asymmetry of x-ray spectrometer maxima arises principally from three-dimensional properties of the ideally two-dimensional focusing arrangement employed, vertical divergence of the primary beam being the fundamental cause. A mathematical formula is developed that expresses the degree of asymmetry as a function of the reflection angle and the geometrical properties of the apparatus. The asymmetrical effects predicted in this way are qualitatively confirmed with experimental measurements. Practical suggestions are made for minimizing asymmetry.

INTRODUCTION

USERS of the Norelco Geiger-counter x-ray spectrometer observe that the diffraction contours exhibit varying degrees of asymmetry. If the contours are manually counted and plotted, the asymmetry increases with decreasing reflection angle. Furthermore, the nature of the asymmetry is always the same, *viz.*, the maximum is broader on the low angle side. This phenomenon is related to the geometry of the instrument in a simple manner. Birks[1] has discussed in a qualitative way certain aspects of contour asymmetry with special reference to the effect upon contour position. The present communication considers the origin and extent of asymmetry in a more quantitative manner and shows how it may be minimized in the practical operation of the instrument.

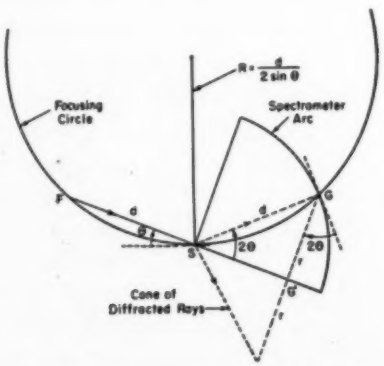


FIG. 1. Two-dimensional geometrical features of the x-ray spectrometer.

In addition to this fundamental type of asymmetry, automatically recorded maxima suffer from a second type of distortion not associated with the geometrical properties of the instrument. This distortion arises from the appreciable discharging time of the tank circuit that supplies the signal to the recorder. The result is that a recorded maximum is broader on the side that is traced last in chronological sequence, the effect being more pronounced as the scanning speed is increased. No further consideration will be extended to this superficial type of asymmetry, it being assumed that for all precise measurements the diffracted quanta are counted and the contours plotted.

SPECTROMETER GEOMETRY

Figure 1 shows the hypothetically two-dimensional focusing arrangement employed in the x-ray spectrometer. Diverging x-rays from the narrow source F irradiate the flat powder sample S, and the diffracted rays from some given reflection are brought to an approximate focus at G, the energy being recorded by a Geiger-Müller tube. In this paper it is assumed that the more efficient spectrometer design, which utilizes a narrow x-ray focal spot as the source of diverging x-rays—thus eliminating a separate "source" slit, is employed. The source F and receiving slit G lie on, and the sample S is tangent to, the hypothetical focusing circle of radius $R=d/2\sin\theta$. In this expression d is both the sample-to-source and sample-to-receiver distance. Suitable apertures are provided to limit the horizontal and vertical divergence of the x-ray beam and to

* Based on material presented before the Fifth Annual Pittsburgh Conference on X-Ray and Electron Diffraction at Pittsburgh, Pennsylvania, November 8, 1947.

[1] L. S. Birks, U. S. Naval Research Laboratory Report No. H-2517, April 20, 1945.

restrict the portion of the diffracted rays received by the Geiger tube.

The fundamental asymmetry with which we are concerned is caused very largely by three-dimensional properties of the *ideally* two-dimensional focusing arrangement described above. Vertical divergence of the primary beam is the fundamental cause. The degree of this divergence is limited by the heights of the x-ray focal spot, beam-defining (x-ray) slit, and receiving slit.

Figure 2 illustrates schematically the extent of the first two vertical effects. F, X, and S are, respectively, the positions of the x-ray source, x-ray slit, and sample; G' represents the projection of the receiving slit on the vertical plane which includes F, X, and S (see also Fig. 1); h_f and h_x denote, respectively, the heights of the x-ray source and x-ray slit.

It is seen that x-rays diverging from the various vertical levels of the source F distribute themselves over a linear interval $H=L+M$ at G'. This means that when a sample is interposed at S, a corresponding multiplicity of diffracted cones is generated, the receiving slit G intercepting small arcs of circular sections of these hollow cones. The centers of these diffraction halos distribute themselves over the vertical interval H at G'. From Fig. 2 it is clear that

$$L/(d+d\cos 2\theta)=h_x/b,$$

and

$$M/(c+d\cos 2\theta)=h_f/b,$$

from which H is given by

$$H=L+M$$
$$=(dh_x/b)(1+\cos 2\theta)+(h_f/b)(c+d\cos 2\theta). \quad (1)$$

In these equations b is the distance from the x-ray source to the x-ray slit, c is the distance from the x-ray slit to the sample, and the other quantities have the meanings presented earlier.

THE ORIGIN OF ASYMMETRY

In Fig. 3 an effort has been made to portray the arrangement of diffraction halos for the particular case of diffraction at an angle $2\theta=10°$. The drawing is made to scale for the following typical values of the various quantities involved in Eq. (1): $b=8.5$ cm, $c=4.5$ cm, $d=13$ cm, $h_x=h_g=0.35$ cm. Each halo has a radius of

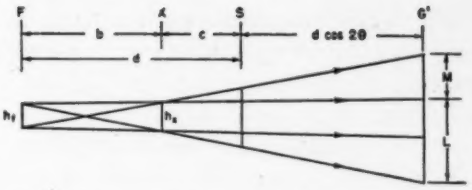


FIG. 2. Vertical divergence of the primary beam in the x-ray spectrometer.

$r=d\sin 2\theta=2.26$ cm (see Fig. 1) and is represented as a line trace of infinitesimal width, although actually it has appreciable width and for most purposes may be assumed to possess an error-curve intensity contour. From a qualitative inspection of Fig. 3 it is plain that, as the receiving slit moves across the family of diffraction halos in the direction of the arrow, the energy measured will rise rapidly to a maximum and fall off much more slowly to a minimum. In short, the plotted maximum will be highly asymmetric with the greater width on the low angle side of the point of maximum intensity, a result in agreement with observation.

The finite width of the x-ray source, which may be assumed to possess an exponential intensity cross section, and other less important two-dimensional geometrical features combine to produce a basic diffraction contour which, at least for particles with a mean dimension larger than 0.10 micron, is approximately an error curve in shape. When a narrow receiving slit is employed, the width of this basic contour at half-maximum

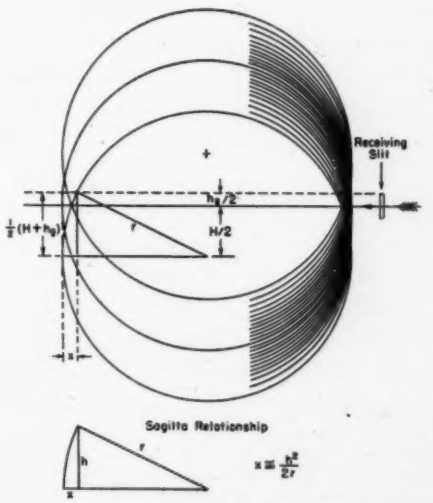


FIG. 3. Multiplicity of diffraction halos for $2\theta=10°$.

intensity is observed to be of the order of $0.20°2\theta$. Figure 4 portrays the synthesis of an asymmetric diffraction contour (dashed line) by the combined action of a large number of such error curves distributed in the manner shown schematically in Fig. 3 for the case $2\theta=10°$.

The degree of asymmetry can be expressed in a semiquantitative manner if it is defined as the total linear or angular range over which an infinitely narrow receiving slit intercepts the centers of the many over-lapping error-curve contours which together build up the final diffraction contour. In Fig. 3 this distance is called x. In terms of an actual asymmetric contour, x is an approximate measure of how much further the contour extends on one side of the point of maximum intensity than it does on the other.

A rather accurate expression for x can be derived by employing the well-known formula which expresses the length of the sagitta in terms of the circle radius r and the ordinate h (see Fig. 3, insert),

$$x \cong h^2/2r, \tag{2}$$

which is correct to within a few percent for $h \ll r$. In the present application h is the quantity $\frac{1}{2}(H+h_g)$ and r is the radius of any diffraction halo. Then we obtain from (2)

$$\begin{aligned} x(\text{cm}) &= [\tfrac{1}{2}(H+h_g)]^2/2r \\ &= (H+h_g)^2/8r. \end{aligned} \tag{3}$$

Since diffraction maxima are plotted in terms of angular units ($°2\theta$), it is also more useful to express the magnitude of the asymmetry in the same angular units rather than in the linear units (cm) of Eq. (3). To effect this transformation it is necessary to take into account the fact that the receiving slit does not move in the plane of Fig. 3 but along the spectrometer arc, which makes an angle 2θ with this plane. From an inspection of Fig. 1 it can be seen that a factor $\cos 2\theta$ must be introduced to allow for this circumstance. When Eq. (3) is modified in this way and r is replaced by $d \sin 2\theta$ and H by its expression given in Eq. (1), we obtain

$$\begin{aligned} x(°2\theta) &= x(\text{cm}) \times (57.3 \cos 2\theta/d) \\ &= [(H+h_g)^2/8d \sin 2\theta] \times [57.3 \cos 2\theta/d] \\ &= (7.16/d^2)[(dh_x/b)(1+\cos 2\theta) \\ &\quad + (h_f/b)(c+d \cos 2\theta)+h_g]^2 \cot 2\theta. \end{aligned} \tag{4}$$

Equation (4) is a general expression for the asymmetry in terms of the geometrical constants of the x-ray spectrometer. If b, c, d, and h_f are assigned their proper respective values for the Norelco spectrometer of 8.5, 4.5, 13.0, and 0.28 cm, this equation becomes

$$\begin{aligned} x(°2\theta) &= 0.0423[1.53(1+\cos 2\theta)h_x \\ &\quad + (0.148+0.429 \cos 2\theta)+h_g]^2 \cot 2\theta. \end{aligned} \tag{5}$$

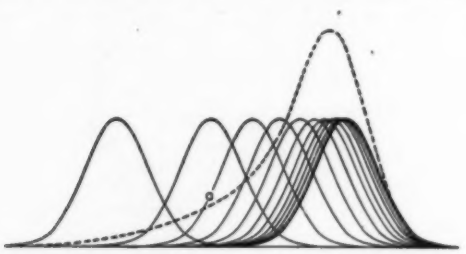


FIG. 4. Synthesis of an asymmetric diffraction contour from an asymmetric distribution of error-curve contours. In plotting the resultant contour the ordinate values have been scaled down by a factor of five. To the left of the point a irregularities arising from the use of a finite number of error-curve contours have been eliminated in drawing the final curve.

COMPARISON WITH EXPERIMENT

Equation (5) predicts (a) that the asymmetry will increase with increasing height of both the x-ray and the receiving slit and that the former will exert the greater effect, and (b) that the asymmetry will increase with decreasing reflection angle, the increase becoming rapid at low angles in accord with the function $\cot 2\theta$. Figure 5 illustrates the effect of slit heights by measurements of the 3.33A quartz maximum, first with the heights of both slits set at 0.20 cm and secondly at 0.50 cm. As expected, the peak is only slightly asymmetric in the former case but decidedly so in the latter case. In Fig. 6 the increase of asymmetry with decreasing reflection angle is strikingly portrayed in the measured contours of the 1.82 and 4.25A quartz maxima, the respective 2θ values being 50.0 and 20.9° for CuKα radiation. Both sets of measurements were made with the heights of both slits fixed at 0.35 cm and all other conditions except angular setting kept constant.

DISCUSSION

Minor factors contributing to contour asymmetry are (a) the use of a flat sample surface tangent to the focusing circle rather than a

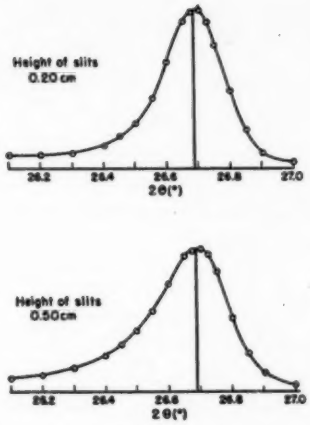


FIG. 5. Observed effect of slit heights upon the symmetry of the 3.33A quartz maximum.

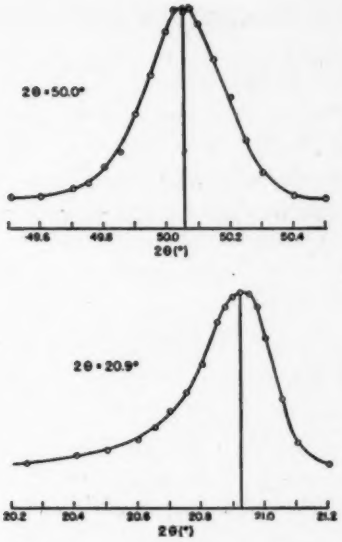


FIG. 6. Observed effect of diffraction angle upon the symmetry of quartz maxima.

sample surface curved to coincide with the focusing circle and (b) penetration of the sample to a finite depth by the incident x-ray beam. Except for sample materials of unusually low absorbing power for x-rays, both of these factors are negligible in comparison with the third-dimensional properties of the apparatus which have been discussed above.

It is evident that asymmetry cannot be present without a certain accompanying degree of widening of the diffraction contour and consequent reduction in resolving power. Hence the present study serves to point out that the use of the smallest practicable slit height has the twofold effect of minimizing asymmetry and decreasing contour width. If the available intensity under these conditions is inadequate, it is best from the standpoint of symmetry to increase the height of the receiving slit but to keep the x-ray slit height as small as possible. Furthermore, if there is special need for very symmetrical maxima and the choice of the particular reflection is not critical, it is best to make use of a reflection at moderately high angles, say $2\theta = 40°$ or more.

In conclusion, it should be mentioned that vertical divergence of the x-ray beam, together with the resulting contour asymmetry, can be largely eliminated by the use of Soller slits. However, this greatly reduces the intensity of the diffracted beam so that provision must be made for either augmenting the intensity of the primary beam or increasing the sensitivity of the receiver or both.

ACKNOWLEDGMENT

The author is grateful to Dr. Harold P. Klug for his interest and encouragement during the course of this study and for helpful criticisms during the preparation of the manuscript.

Theory of the Positive Column in an Annular Discharge Tube

JAROSLAV PACHNER
Prague, Czechoslovakia
(Received June 14, 1948)

On the ground of Schottky's diffusion theory of the positive column, the distribution of the electron and ion concentration and of the current density in an annular discharge tube are calculated.

INTRODUCTION

IN 1924 W. Schottky[1] built up his diffusion theory of the positive column and computed on this ground the distribution of the electron and ion concentration in an infinite cylindrical discharge tube. Five years later, in 1929 L. Tonks and I. Langmuir[2] proved that, in the case of short mean free paths and ion generation proportional to the electron density, even his general theory of the plasma of an arc gives the same distribution of the electron and ion concentration. The assumptions of this general theory are much closer to reality, but on the other hand the great advantage of the diffusion theory consists in that it uses a differential equation which permits one to calculate the electron and ion concentration not only in the steady-state but also during the de-ionization of the plasma.[3] Because its results agree with that found experimentally, it is used to this day for the study of electric discharges at low pressures.

In order to ascertain the influence of de-ionization grids, E. Spenke and M. Steenbeck[4] extended Schottky's diffusion theory to any cross section of an infinite straight positive column.

Similarly when the importance of annular discharges in technical applications became evident, Th. Wasserrab[5] tried to solve this mathematically difficult problem. He simplified it by the assumption that the discharge space is bounded by two coaxial infinite cylinders between which the positive column burns annularly.

In this article the same problem is solved exactly: the positive column burns in an annular discharge tube without electrodes.

THE COMPUTATION OF THE ELECTRON AND ION CONCENTRATION

The fundamental differential equation of the steady-state positive column in Steenbeck's formulation is

$$\Delta n+(1/D_a\tau)n=0, \tag{1}$$

where n is the electron concentration, equal to the ion concentration, D_a the ambipolar diffusion coefficient, τ the mean life-time of an ion, and Δ Laplace's differential operator.

The boundary conditions in the diffusion theory are: the electron and ion concentration is finite in the discharge space, and it is zero at the tube walls.

Under these conditions it is suitable to introduce toroidal coordinates defined by the following relations:[6]

$$x=c\frac{\sinh\xi}{\cosh\xi-\cos\eta}\cos\varphi,$$

$$y=c\frac{\sinh\xi}{\cosh\xi-\cos\eta}\sin\varphi,$$

$$z=c\frac{\sin\eta}{\cosh\xi-\cos\eta},$$

[1] W. Schottky, "Diffusionstheorie der positiven Säule," Phys. Zeits. **25**, 342, 635 (1924).

[2] L. Tonks and I. Langmuir, "A general theory of the plasma of an arc," Phys. Rev. **34**, 635 (1929).

[3] E. Spenke and M. Steenbeck, "Zur Theorie der positiven Säule bei beliebigen Querschnittsformen," Wiss. Veröff. Siemens **XV**, 18 (1936). Th. Wasserrab, "Zur Beschreibung des Entionisierungsvorganges von Gasentladungen," Wiss. Veröff. Siemens **XIX**, 1 (1940). W. Schottky and J. v. Issendorff, "Quasineutrale elektrische Diffusion im ruhenden und strömenden Gas," Zeits. f. Physik **31**, 163 (1925). J. Pachner, "Theory of the de-ionization in flowing gases." J. Pachner, "Transient phenomena in the positive column." (These two articles will be published in Czech language in Elektrotechnický Obzor, Prague 1948).

[4] E. Spenke and M. Steenbeck, reference 3.

[5] Th. Wasserrab, "Die positive Säule mit gekrümmter Entladungsbahn," Zeits. f. Physik. **121**, 54 (1943).

[6] W. Magnus and F. Oberhettinger, *Formeln und Sätze für die speziellen Funktionen der mathematischen Physik.* (Berlin, 1943), p. 151.

where

$$0 \leqq \xi \leqq \infty, \quad -\pi \leqq \eta \leqq +\pi, \quad 0 \leqq \varphi \leqq 2\pi.$$

The relation

$$\xi = \xi_0 = \text{const.} \tag{2a}$$

determines the surface of an annuloid, the equation of which in Cartesian coordinates is

$$((x^2+y^2)^{\frac{1}{2}} - c \operatorname{ctgh}\xi_0)^2 + z^2 = \left(\frac{c}{\sinh\xi_0}\right)^2. \tag{2b}$$

The cross section of this annuloid with the plane $y=0$ is shown in Fig. 1. We see from it and from Eq. (2b) that the relation between the constants ξ_0, c on the one hand and the radii R_0, r_0 on the other hand are given by

$$r_0 = \frac{c}{\sinh\xi_0},$$

$$R_0 = c \operatorname{ctgh}\xi_0.$$

Hence,

$$\left.\begin{aligned} c &= (R_0^2 - r_0^2)^{\frac{1}{2}}, \\ \cosh\xi_0 &= \frac{R_0}{r_0}. \end{aligned}\right\} \tag{3}$$

Assuming rotational symmetry about the Z axis

$$\frac{\partial n}{\partial \varphi} = 0,$$

the differential Eq. (1) is expressed in toroidal coordinates as follows:

$$\frac{\partial}{\partial\xi}\left[g(\xi,\eta)\frac{\partial n}{\partial\xi}\right] + \frac{\partial}{\partial\eta}\left[g(\xi,\eta)\frac{\partial n}{\partial\eta}\right] + g(\xi,\eta)\frac{k}{(\cosh\xi - \cos\eta)^2} n = 0, \tag{4}$$

where

$$\left.\begin{aligned} g(\xi,\eta) &= \frac{\sinh\xi}{\cosh\xi - \cos\eta}, \\ k &= \frac{c^2}{D_a\tau}. \end{aligned}\right\} \tag{5}$$

The boundary conditions are given by

$$\left.\begin{aligned} &n = 0 \quad \text{at} \quad \xi = \xi_0, \\ \text{and} \quad &n \text{ finite} \quad \text{at} \quad \xi = \infty. \end{aligned}\right\} \tag{6}$$

It is not possible to separate the variables ξ and η in Eq. (4).[7] We use therefore the substitution

$$n(\xi,\eta) = h(\xi,\eta)u(\xi,\eta). \tag{7}$$

The function $h(\xi,\eta)$ is determined by the condition that the coefficients at $\partial u/\partial\xi$ and $\partial u/\partial\eta$ become zero. We find

$$h(\xi,\eta) = \frac{1}{(g(\xi,\eta))^{\frac{1}{2}}} = \left(\frac{\cosh\xi - \cos\eta}{\sinh\xi}\right)^{\frac{1}{2}}. \tag{8}$$

By the substitution of (7) Eq. (4) is changed into the form

$$\frac{\partial^2 u}{\partial\xi^2} + \frac{\partial^2 u}{\partial\eta^2} + f(\xi,\eta)u = 0, \tag{9}$$

$$f(\xi,\eta) = \frac{1}{(\cosh\xi - \cos\eta)^2} \times \left\{(k+\tfrac{1}{4}) - \frac{1}{2}\cdot\frac{\cosh\xi\cos\eta - \dfrac{1+\cos^2\eta}{2}}{\sinh^2\xi}\right\}. \tag{10}$$

Now we shall inspect the behavior of the solution of (9) for large values of ξ. In this case we

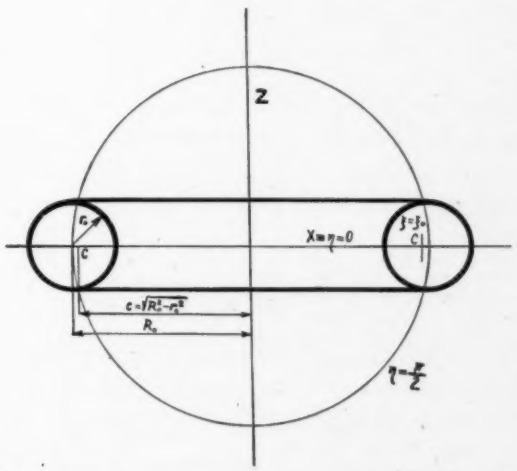


FIG. 1. The cross section of an annular discharge tube.

[7] Reference 6, p. 158.

obtain the following expression for the function $f(\xi, \eta)$:

$$\lim_{\xi \to \infty} f(\xi, \eta) = (4k+1)e^{-2\xi}.$$

Putting

$$\lim_{\xi \to \infty} u(\xi, \eta) = u_0,$$

we have

$$\frac{\partial^2 u_0}{\partial \xi^2} + \frac{\partial^2 u_0}{\partial \eta^2} + (4k+1)e^{-2\xi}u_0 = 0.$$

If we assume the solution of this equation to be of the form

$$u_0(\xi, \eta) = \cos\mu\eta \cdot v(\xi),$$

we obtain an ordinary differential equation

$$\frac{d^2 v}{d\xi^2} + [(4k+1)e^{-2\xi} - \mu^2]v = 0,$$

the integral of which is the Bessel function

$$v(\xi) = Z_\mu(e^{-\xi}(4k+1)^{\frac{1}{2}}).$$

With regard to these relations, only the Bessel function of the first kind and of the order zero satisfies the boundary conditions (6). We obtain thus as the asymptotic solution of (9)

$$u_0 = J_0(e^{-\xi}(4k+1)^{\frac{1}{2}}), \tag{11}$$

where $\mu = 0$. Its eigenvalue is

$$(4k+1)^{\frac{1}{2}} = x_1 e^{\xi_0}. \tag{12}$$

Here x_1 signifies the first root of the equation $J_0(x) = 0$, the value of which is given according to E. Jahnke and F. Emde[8] by

$$x_1 = 2.405. \tag{13}$$

Putting (12) and (13) into (10) we find that for $R_0/r_0 = \cosh\xi_0 \geqq 4$ the value of the second term in the large bracket is $\leqq 0.1$ percent of the first one. Therefore we may write with very good accuracy

$$f(\xi, \eta) \cong \frac{k+\frac{1}{4}}{(\cosh\xi - \cos\eta)^2}. \tag{14}$$

To obtain relations that are known from the theory of Bessel functions, we choose a new variable ρ defined by

$$\rho = e^{\xi_0 - \xi} = e^{-\xi}/\epsilon. \tag{15}$$

[8] E. Jahnke and F. Emde, *Funktionentafeln. 3. Aufl.* (Berlin, 1938), p. 166.

Using (14) we obtain by this substitution from (9)

$$\rho^2\frac{\partial^2 u}{\partial \rho^2} + \rho\frac{\partial u}{\partial \rho} + \frac{\partial^2 u}{\partial \eta^2} + \lambda\frac{\rho^2}{(1 - 2\epsilon\rho\cos\eta + \epsilon^2\rho^2)^2}u = 0, \tag{16}$$

where

$$\lambda = \epsilon^2(4k+1). \tag{17}$$

The boundary conditions follow from (6),

$$\left.\begin{aligned} &u(0, \eta) \text{ finite,} \\ &u(1, \eta) = 0. \end{aligned}\right\} \tag{18}$$

By means of the binomial formula we transform Eq. (16) into the form

$$\rho^2\frac{\partial^2 u}{\partial \rho^2} + \rho\frac{\partial u}{\partial \rho} + \frac{\partial^2 u}{\partial \eta^2} + \lambda\rho^2 u = -\lambda\rho^2 u[4\epsilon\rho\cos\eta + \epsilon^2(\cdots) + \cdots]. \tag{19}$$

If $\epsilon = 0$, we have the equation

$$\rho^2\frac{\partial^2 u_0}{\partial \rho^2} + \rho\frac{\partial u_0}{\partial \rho} + \frac{\partial^2 u_0}{\partial \eta^2} + \lambda_0\rho^2 u_0 = 0, \tag{20}$$

the eigenfunction of which is identical with the asymptotic solution (11)

$$u_0 = J_0(x_1\rho). \tag{21}$$

Its eigenvalue is therefore,

$$\lambda_0 = x_1^2. \tag{22}$$

The parameter ϵ has a small value ($\epsilon \cong \frac{1}{8}$ for $R_0/r_0 = 4$). In first approximation we neglect therefore all the terms in ϵ^2, ϵ^3, $\cdots$ and assume the solution of (19) to be of the form

$$u = u_0 + \epsilon\bar{u}, \tag{23}$$

with eigenvalue

$$\lambda = \lambda_0 + \epsilon\bar{\lambda}. \tag{24}$$

By known perturbation methods[9] we obtain from (19)

$$\left(\rho^2\frac{\partial^2 u_0}{\partial \rho^2} + \rho\frac{\partial u_0}{\partial \rho} + \frac{\partial^2 u_0}{\partial \eta^2} + \lambda_0\rho^2 u_0\right) + \epsilon\left(\rho^2\frac{\partial^2 \bar{u}}{\partial \rho^2} + \rho\frac{\partial \bar{u}}{\partial \rho} + \frac{\partial^2 \bar{u}}{\partial \eta^2} + \lambda_0\rho^2\bar{u}\right) = -\epsilon\rho^2(\bar{\lambda}u_0 + 4\lambda_0\rho u_0\cos\eta). \tag{25}$$

[9] R. Courant and D. Hilbert, *Methoden der mathematischen Physik. Bd.* I. *2. Aufl.* (Berlin 1931), p. 296.

Also here the terms in ϵ^2 and ϵ^3 are neglected. It follows from (20) that the sum of the terms in the first bracket in (25) is zero identically. Because it is possible to expand the term $\rho u_0 \cos\eta$ in a series of Bessel functions[10]

$$\rho u_0 \cos\eta = \cos\eta \sum_{\nu=1}^{\infty} a_\nu J_1(y_1\rho), \tag{26}$$

where y_ν denotes the ν-root of the equation $J_1(y)=0$, we can assume the solution of (25) to be of the form

$$\bar{u} = \cos\eta \sum_{\nu=1}^{\infty} c_\nu J_1(y_\nu\rho). \tag{27}$$

If we put (26) and (27) into (25), we obtain by means of (22)

$$\rho^2 \cos\eta \sum_{\nu=1}^{\infty} (-y_\nu^2+x_1^2)c_\nu J_1(y_1\rho)$$

$$= -\rho^2[\bar{\lambda}J_0(\rho x_1)+4x_1^2 \cos\eta \sum_{\nu=1}^{\infty} a_\nu J_1(y_\nu\rho)]. \tag{28}$$

This equation can be satisfied only if

$$\bar{\lambda}=0, \tag{29}$$

and

$$c_\nu = \frac{4x_1^2}{y_\nu^2-x_1^2}a_\nu. \tag{30}$$

By multiplication of Eq. (26) by $J_1(y_\nu\rho)\rho d\rho$ and integration we compute the values of the coefficients a_ν:

$$a_\nu = \frac{\int_0^1 \rho^2 J_0(x_1\rho)J_1(y_\nu\rho)d\rho}{\int_0^1 J_1^2(y_\nu\rho)\rho d\rho}. \tag{31}$$

The integrals in the numerator of (31) were evaluated by Simpson's method:

$$\int_0^1 \rho^2 J_0(x_1\rho)J_1(y_\nu\rho)d\rho = \begin{cases} +0.0488 & \text{for } \nu=1 \\ -0.0028 & \nu=2 \\ +0.0114 & \nu=3. \end{cases}$$

As the values of y_ν are[11]

$$y_1=3.832, \quad y_2=7.016, \quad y_3=10.174.$$

and

$$\int_0^1 J_1^2(y_\nu\rho)\rho d\rho = \begin{cases} +0.0811 & \text{for } \nu=1 \\ +0.0450 & \nu=2 \\ +0.0312 & \nu=3, \end{cases}$$

we have

$$\left.\begin{aligned} c_1 &= +1.56 \\ c_2 &= -0.033 \\ c_3 &= +0.086. \end{aligned}\right\} \tag{32}$$

We see that we can neglect in the series (27) the terms with $\nu \geqq 2$. Putting the first term into (23) and this relation with (8) into (7) we obtain the resultant equation

$$n(\xi,\eta) = N_0[J_0(x_1e^{\xi_0-\xi})+1.56e^{-\xi_0}J_1(y_1e^{\xi_0-\xi})\cos\eta] \times\left(\frac{\cosh\xi-\cos\eta}{\sinh\xi}\right)^{\frac{1}{2}}, \tag{33}$$

which gives the distribution of the electron and ion concentration in an annular discharge tube with an error amounting to a few percent. N_0 denotes the concentration being at $\xi=\infty$.

The maximum concentration is at $\sin\eta=0$, i.e., in the plane $z=0$, for

$$\left(\frac{\partial n}{\partial \eta}\right)_{\sin\eta=0} = 0.$$

To find the ξ-coordinate of the maximum, we use the substitution (15) and expand the function (33) in a MacLaurin series. If $\cos\eta=1$, we have

$$n = N_0[1+1.89\epsilon\rho-\tfrac{1}{2}(2.89-8.89\epsilon^2)\rho^2+\cdots]$$
$$= N_0[1+1.89e^{-\xi}-\tfrac{1}{2}(2.89\epsilon^{-2}-8.89)e^{-2\xi}+\cdots].$$

Hence,

$$e^{-\xi_{\max}} = \frac{1.98\epsilon^2}{2.89-8.89\epsilon^2}, \tag{34a}$$

or in Cartesian coordinates

$$((x^2+y^2)^{\frac{1}{2}})_{\max} = cg(\xi_{\max},0)$$
$$\cong R_0\left(1-0.15\frac{r_0^2}{R_0^2}\right). \tag{34b}$$

The mean life-time of an ion is determined by

$$\tau = \frac{c^2}{D_a k} \cong \left(1-\frac{r_0^2}{R_0^2}\right)\frac{r_0^2}{D_a x_1^2}. \tag{35}$$

We see that it is shorter than in an infinite cylinder, but the difference can be neglected.

[10] Reference 9, p. 261.
[11] Reference 8, p. 166.

THE COMPUTATION OF THE ELECTRIC CURRENT DENSITY DISTRIBUTION

In a straight discharge tube the electric current density is proportional to the electron concentration, because the electric field intensity is constant across the cross section of the discharge tube. In an annular discharge the electric field intensity changes in values across the cross section and we have therefore to compute it.

We assume that only a little part of the whole magnetic flux goes through the discharge tube. It follows from Maxwell's equation

$$\oint E_s ds = -\frac{\partial}{\partial t}\int B_n dS,$$

that the intensity of the electric field, the direction of which is identical with that of the variable φ, is indirectly proportional to the radius

$$(x^2+y^2)^{\frac{1}{2}} = cg(\xi, \eta).$$

We have thus

$$E \sim \frac{1}{(x^2+y^2)^{\frac{1}{2}}} = \frac{1}{cg(\xi, \eta)}.$$

The electric current density in a plasma is given by the known relation

$$j = \epsilon n b^- E,$$

where ϵ denotes the electron charge and b^- its mobility.

From these two relations and Eq. (33) we obtain

$$j = j_0[J_0(x_1 e^{\xi_0-\xi}) + 1.56e^{-\xi_0}J_1(y_1 e^{\xi_0-\xi})\cos\eta] \times\left(\frac{\cosh\xi-\cos\eta}{\sinh\xi}\right)^{\frac{1}{2}}, \quad (36)$$

where j_0 signifies the current density at $\xi=\infty$.

Also here the maximum of the function (36) is at $\sin\eta=0$, i.e., in the plane $z=0$, because

$$\left(\frac{\partial n}{\partial \eta}\right)_{\sin\eta=0} = 0.$$

Similarly, to find the ξ-coordinate of the maximal current density, we use the substitution (15) and expand the function (36) in a MacLaurin series. If $\cos\eta=-1$, we find

$$n = N_0[1+0.02\epsilon\rho-\tfrac{1}{2}(2.89-20.9\epsilon^2)\rho^2+\cdots] = N_0[1+0.02e^{-\xi}-\tfrac{1}{2}(2.89\epsilon^{-2}-20.9)e^{-2\xi}+\cdots].$$

Hence,

$$e^{-\xi_{\max}} = \frac{0.02\epsilon^2}{2.89-20.9\epsilon^2} \cong 0, \quad (37a)$$

or in Cartesian coordinates

$$((x^2+y^2)^{\frac{1}{2}})_{\max} = cg(\xi_{\max}, \pi) \cong (R_0^2-r_0^2)^{\frac{1}{2}} \cong R_0\left(1-\frac{1}{2}\frac{r_0^2}{R_0^2}\right). \quad (37b)$$

SUMMARY

1. The distribution of the electron and ion concentration in an annular discharge tube is given in first approximation by Eq. (33).

2. The maximum concentration is in the plane $z=0$ at a circle the radius of which is greater than $c=(R_0^2-r_0^2)^{\frac{1}{2}}$ and given by Eq. (34).

3. The distribution of the electric current density is directly proportional to the electron concentration and indirectly proportional to the distance of this point from the Z axis.

4. It is given in first approximation by Eq. (36).

5. The maximum current density is in the plane $z=0$ at a circle the radius of which equals very approximately to the radius $c=(R_0^2-r_0^2)^{\frac{1}{2}}$.

6. The greater is the value of R_0/r_0, the less the radius of maximum concentration and of maximum current density differs from that one which goes through the center of the annular discharge tube and,

7. The less the distribution of the electron and ion concentration differs from that in an infinite cylinder, i.e., from the distribution given by the Bessel function.

The quantitative results of this theory will agree with that found experimentally only under assumption of the validity of the Schottky's diffusion theory. If the electric current intensity in an annular discharge tube reaches some hundreds of amperes, the distribution will be changed. M. Steenbeck[12] proved that in consequence of the "electron gas" pressure it will have a flatter maximum and will fall more steeply to zero at the walls. But also in such cases the distribution will remain qualitatively the same as expressed sub 2, 3, 5, and 6.

[12] M. Steenbeck, "Theoretische und experimentelle Untersuchung über den Einfluss des Elektronenpartialsdrucks," Wiss. Veröff. Siemens **XVIII**, p. 318 (1939).

Electron Diffraction Study of Particle Size in Thin Bi Films Deposited by Evaporation in Vacuum

C. T. KEOGH AND A. H. WEBER
St. Louis University, St. Louis, Missouri
(Received June 21, 1948)

Electron diffraction patterns of thin Bi films were studied by the method of microphotometer density curve analysis to determine changes of the film structure with aging and with increase in thickness of the films. All Bi films were deposited on Formvar by evaporation in vacuum. Analysis of the Bi diffraction patterns by comparison with the Formvar pattern as a reference standard indicate that the very thin films (about 20 atom layers or less in thickness) while initially composed of small crystals grow rapidly upon aging into larger crystals. The very thinnest film (about 5 atom layers thick) analyzed exhibited an initial increase in particle size followed by a decrease upon aging, a somewhat anomalous result requiring further confirmation.

Increase in particle size with increase in thickness of the Bi films is observed also, although the results of these latter experiments are somewhat irregular.

A VARIETY of photoelectric and photoconductance experiments[1,2] with thin Bi films evaporated on Pyrex in vacuum have indicated mobility of the deposited Bi atoms over the film surface so that the film finally is composed of discrete aggregates of Bi. This paper describes an investigation by electron diffraction of the change of particle size in thin Bi films deposited by evaporation in vacuum during aging of films of constant average thickness and during increase in thickness of such films. The electron diffraction camera used follows the design of J. E. Ruedy[3] and in addition includes the photographic plate holder designed by A. H. Weber and C. H. Dahm,[4] the electron gun modifications described in Fig. 1 and the accompanying legend, and modifications to provide for evaporation of metallic Bi within the diffraction apparatus.

EXPERIMENTAL

Several experimenters, notably L. H. Germer,[5] have used Formvar as a substrate in electron diffraction experiments. The Formvar[6] films used in the present experiments were prepared as follows. A 0.1 percent solution of Formvar in ethylene dichloride (100 mg Formvar type "*E*" to 100 ml of ethylene dichloride) was prepared. A culture dish approximately 2 in. in diameter and 1 in. deep was filled with distilled water to within $\frac{1}{8}$ in. of the top. A copper screen (about 17 mesh) of suitable size for mounting was hammered to flatness and placed in the water-filled dish. Then with a pipette a drop of the Formvar solution was placed carefully on the water surface. After about 10 sec. a second drop was placed carefully on the water surface directly above the copper screen lying at the bottom of the dish. After the ethylene dichloride had evaporated almost completely, the remaining Formvar film was caught on the copper screen by removing the screen from the liquid. A smooth surface of Formvar will result if the screen is tilted slightly when the film is being lifted off the liquid surface. The screen with the wet Formvar film then was placed in a clean, covered drying dish containing no drying agent. After about 24 hours the film may be examined with a microscope using reflected light to locate the thinner sections. Even though a somewhat low intensity and low velocity electron beam (approximately 0.07 μamp. at 32.8 kv in the present investigation) is used, exposures of less than 1 sec. show the diffuse ring pattern of Formvar.

The electron diffraction patterns were recorded on Eastman Kodak spectroscopic plates type IV-O. This type of plate was chosen after experi-

[1] A. H. Weber and D. F. O'Brien, Phys. Rev. **60**, 574 (1941).
[2] A. H. Weber and L. W. Friedrich, Phys. Rev. **66**, 248 (1944).
[3] J. E. Ruedy, Rev. Sci. Inst. **11**, 292 (1940).
[4] A. H. Weber and C. H. Dahm, Rev. Sci. Inst. **16**, 141 (1945).
[5] L. H. Germer, Phys. Rev. **56**, 58 (1939).
[6] The Formvar, type "*E*", was furnished by V. J. Schaefer of the General Electric Research Laboratory, Schenectady, New York. The help of V. J. Schaefer in the preparation of suitable Formvar films is acknowledged with gratitude.

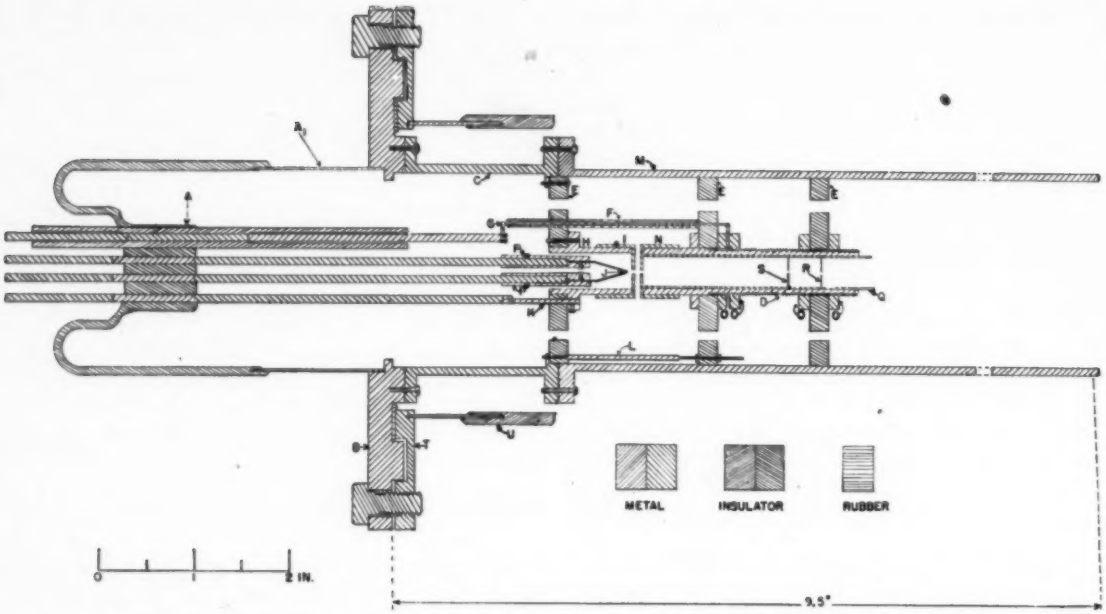


FIG. 1. Electron gun assembly. A, 4-wire glass press with first anode lead insulated with glass tubing; B, brass ring soft soldered to Kovar tube A_1 and bolted to the diffraction apparatus gasket ring T; C, brass tube bolted to B with six brass machine screws at 60° around circumference; D, outer threaded brass tube of first anode assembly; E, perforated baked Lavite insulating and alignment disks; F, glass tube insulating first anode connection; G, 0.05-in. nickel wire spot welded to first anode lead and making connection to first anode at knurled nuts O; H, brass grid cylinder bolted to Lavite insulator E and connected to grid lead by nickel lead K; I, grid cap fastened with set screw and having 0.03-in. hole; J, 0.005-in. thoriated W wire filament fastened by set screws to filament leads; L, one of three threaded brass rods screwed into C and holding in place the insulator E of first anode assembly—thus keeping the first anode in alignment; M, second anode consisting of nickel-plated brass tube of polished inside surface; N, first anode cap (with 0.055-in. hole) which holds in place the inner first anode nickel-plated and polished brass tubes Q (3 in number) and the diaphragms R and S; O, knurled nickel-plated brass nuts holding the insulators E in place; P, soft iron filament connectors with set screws; R, diaphragm with 0.03-in. hole; S, diaphragm with 0.02-in. hole; U, glass high voltage insulator sealed to Kovar ring.

menting with Lantern Slide plates and Eastman process plates to determine a plate suitable in speed, contrast, and resolution for electron diffraction recording. The procedure followed in evaporating Bi has been described[7] and the calculation of its deposition rate follows from the measured[8] vapor pressure of Bi at the evaporation temperature used in the present experiments. Film thicknesses then are calculated by the product of deposition rate (in atom layer min.$^{-1}$) and evaporation time (in min.).

Microphotometer density curves (Figs. 2 and 3 are examples) of the photographic negatives were recorded to determine changes in particle size in the Bi films. In constructing these curves it was assumed, following Germer and White[9] who used Eastman process plates, that an approximate linear relationship exists between photographic density and incident electron beam intensity. Comparatively high amplification (the largest microphotometer galvanometer deflection used was 50 cm and the time to scan a 2-in. length plate was 25 min.) was employed to measure changes in light intensity as the scanning beam passed through various rings of the diffraction pattern.

For Figs. 2 and 3 the corrected shapes of the (202) and (200) Bi rings, shown by the inset curves, were obtained[10] by dividing the observed intensities (that is, the difference between corresponding points on the dashed "background" curve and the solid microphotometer trace) by the factor f_e^2/R^2 (f_e is the Bi atomic scattering factor for electrons, R is the ring radius) and plotting the resulting values on an arbitrary ordinate scale marked "Corrected density." In

[7] A. H. Weber and L. J. Eisele, Phys. Rev. 60, 570 (1941).
[8] A. H. Weber and G. Platenberg, Phys. Rev. 69, 649 (1946).
[9] L. H. Germer and A. H. White, Phys. Rev. 60, 447 (1941).
[10] This procedure is adopted from L. H. Germer and A. H. White, reference 9.

turn $f_e=(Z-f)/R^2$, where Z is 81 and values for f (f is the Bi atomic scattering factor for x-rays) are from tables in Buerger.[11]

In this paper all data has been determined to three significant figures; hence in the present work film thicknesses are stated usually in three figures as a convenience [actually (see reference 8) the third figure is doubtful].

RESULTS

In general the experimental results, Figs. 4 and 5, show variation of certain Bi pattern ring widths with aging time and with film thickness. Changes in film particle size may be determined by a direct measurement of diffraction ring width

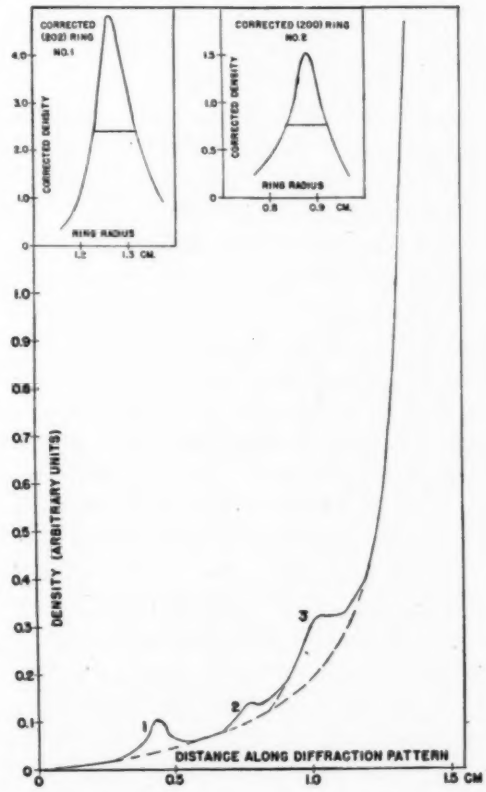


FIG. 2. Microphotometer density curve of Bi film (about 17.8 atom layers thick) deposited on Formvar in vacuum at room temperature and aged six hours. 1, (202) Bi ring trace; 2, (200) ring trace; 3, Formvar trace. Pattern exposure time, 6 sec.; accelerating potential, 32.8×10^3 volts; electron beam current, 0.07 μamp. approximately; maximum microphotometer galvanometer deflection, 28 cm; time required to record the trace, 25 min. Dashed lines represent the estimated background.

[11] M. J. Buerger, *X-Ray Crystallography* (John Wiley & Sons, Inc., New York, 1942), p. 52, Fig. 31. Values of f for Bi were obtained by interpolation.

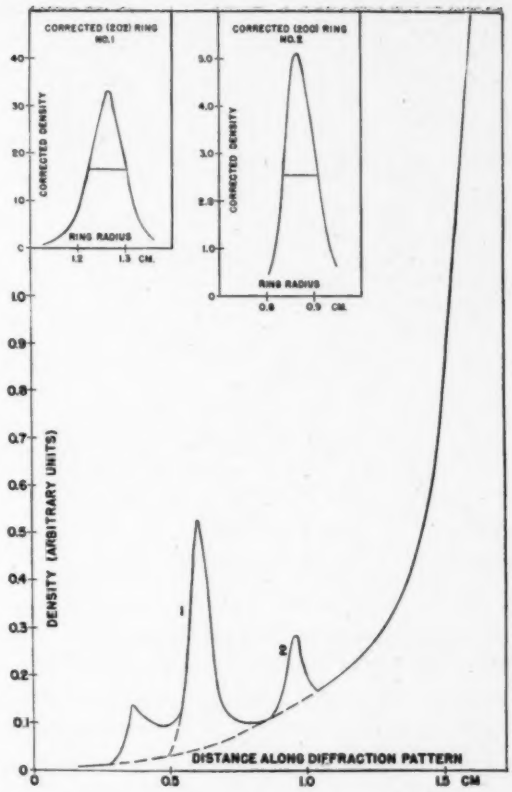


FIG. 3. Microphotometer density curve of Bi film (about 172 atom layers) deposited on Formvar in vacuum at room temperature. 1, (202) ring trace; 2, (200) ring trace. Bi film was not aged, electron diffraction pattern was obtained as soon as Bi film was about 172 atom layers thick; Formvar ring does not appear because the Bi which is quite thick is dominant diffracting material. Pattern exposure time, 9 sec.; accelerating potential, 32.8×10^3 volts; electron beam current, 0.07 μamp. approximately. Dashed lines represent the estimated background. The third peak, not numbered, is from another ring not of present interest.

or by measurement of ring width changes as compared to some ring width which remains substantially unchanged. The size[12] of the diffracting particles may be calculated from the Scherrer formula[13] in the form[14]

$$C=L\lambda/\Delta R, \qquad (1)$$

where C is the particle size, ΔR is width of the diffraction ring, L is the specimen to photographic plate distance, λ is the deBroglie electron wave-length.

[12] By particle size is meant the average linear dimension of the crystallites producing a diffraction ring; cf. L. H. Germer, Phys. Rev. **56**, 58 (1939).

[13] C. Brill, Zeits. f. Krist. **87**, 275 (1934).

[14] This is the form used for experiments with fast electrons.

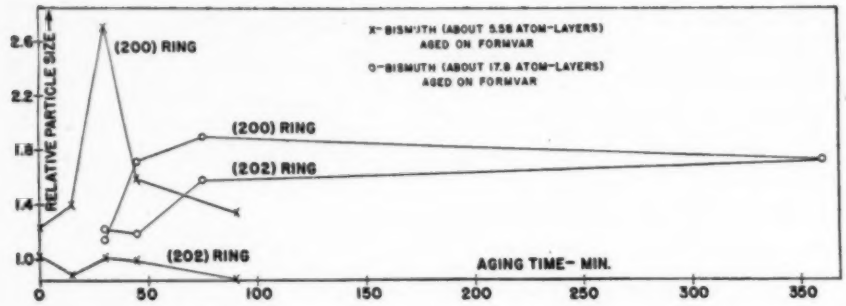


FIG. 4. Relative particle size *vs.* aging time at room temperature in vacuum for the (200) and (202) Bi rings of the 5.58 and 17.8 atom layers thick films. The numerical values along the ordinate are the ratio of width at half-maximum of the Formvar ring to width at half-maximum of the Bi ring in question. Increase in ordinate corresponds to increase in sharpness of the diffraction pattern rings indicating an increase in particle size. Diffraction patterns for aging times less than 30 min. were made for the 17.8 atom layer film but were not suitable for microphotometer analysis. The 17.8 atom layer film is the same as that of Fig. 2.

Equation (1) does not include the instrument line width; hence variations of ring width caused by wandering of the electron beam, lack of uniformity in beam structure, and variations of beam cross section have to be taken into account in applying the equation.

If the proper exposure time is used, both the Formvar and the Bi film diffraction patterns will appear and the Formvar diffraction pattern may be used as a reference. (See Fig. 6.) In the present investigation no evidence of changes in Formvar and its diffraction patterns with age was found. Hence, where possible the Formvar electron diffraction pattern was used as a reference standard. For example, the numbers along the ordinate scale of Fig. 4 are values of the ratio of the width at half-maximum of the Formvar ring to the width at half-maximum of the Bi ring in question (both half-widths are obtained from corrected density curves such as the insets on Figs. 2 and 3). If the numerator of this ratio remains constant as we have found, the ratio is directly proportional to the Bi particle size because of Eq. (1). Hence the ordinate of Fig. 4 is marked as shown with relative particle size increasing as ordinate values increase. It is to be noted that if the electron beam wandered, became larger in cross section, or changed in intensity, the width at half-maximum of the Formvar ring pattern would vary likewise and hence the ratio just mentioned should be a reliable means of studying relative changes of the Bi pattern ring widths.

For the experiments on the variation of particle size with Bi film thickness, the method of the last paragraph is not applicable since the measured density of the Formvar ring pattern varies with the amount of deposited Bi. Hence for these experiments, see Fig. 5, the widths at half-maximum of the Bi ring in question are plotted as ordinates *decreasing* in the positive direction so that relative particle size could be marked *increasing* as shown.

Figure 4 shows the change in particle size of two Bi films of different thicknesses as they were aged immediately after formation in the evacuated electron diffraction camera. It is seen that both films show rapid changes in particle size for the period 50 to 75 minutes after the film was formed. This is to be expected since mobility of condensed surface atoms, a process which should be particularly active immediately following condensation of the vapor into a film, is a well established phenomenon in thin metallic films. L. H. Germer[5] has shown by electron diffraction studies of many thin films (not including Bi, however) that the average crystal size perpendicular to the film surface is usually much larger than the mean film thickness, which result he interprets as caused by the tendency of condensed atoms or molecules to move about on the surface of the substrate until fairly large three-dimensional crystal agglomerates are formed. Photoelectric and photo-conductance experiments[1,2] of one of the present authors with thin films of Bi establish

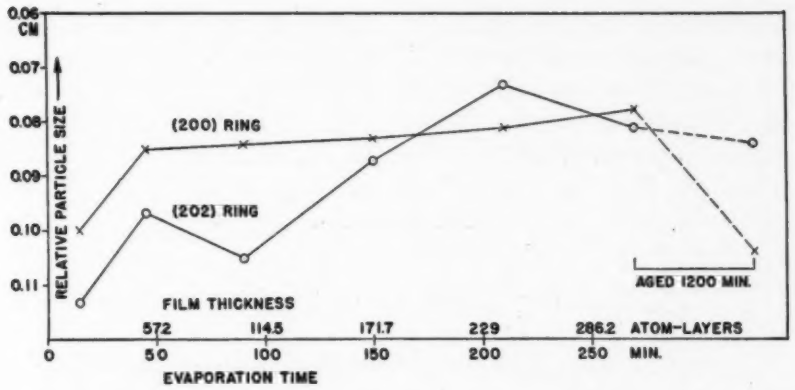


FIG. 5. Relative particle size *vs.* film thickness during continuous Bi deposition in vacuum at room temperature. The numerical values along the ordinate are the widths at half-maximum of the Bi ring in question. Decrease in values of these widths corresponds to *increase* in sharpness of the diffraction rings indicating an increase in particle size. The dashed portions of the curves indicate that the film after being aged 1200 min. (with no increase in Bi film thickness) produced a somewhat more diffuse pattern. The numerical values of film thickness in atom layers are given to four figures to correspond to the exact evaporation times from which they were calculated as explained in the text.

the same conclusion. Hence the curves of Fig. 4 may be useful in indicating some of the characteristics of surface mobility of the atoms.

For the thicker (about 17.8 atom layers) of the two Bi films, a continued *growth* of particle size to an equilibrium value is observed. This indicates a normal process of agglomeration of the condensed metal into somewhat larger particle sizes. The thinner (about 5.58 atom layers) film exhibits more unusual behavior in that a *decrease* in particle size is observed for both the (200) and (202) rings even though the (200) ring showed an initial (i.e., during the first 30 min. of aging) and rapid growth of particle size. It is considered that more data are necessary to establish this phenomenon of growth followed by decay of particle size (to obtain accurate data of this kind is difficult and quite laborious). However, the decay of particle size after the first some 50 min. of aging time would indicate that in very thin films (about 5 atom layers, or less, in thickness) the process of agglomeration of atoms into larger groups does not take place.

Figure 5 shows the particle size changes which occur with varying amounts of deposited Bi. Although the results are somewhat irregular, there are two consistent features evident: (1) initial (during first 50-min. evaporation time) increases of particle size with increasing film thickness, and (2) a saturation particle size which does not vary much with film thickness. The latter result is shown directly by the curve for the (200) ring; it would be shown by a smooth curve ap-

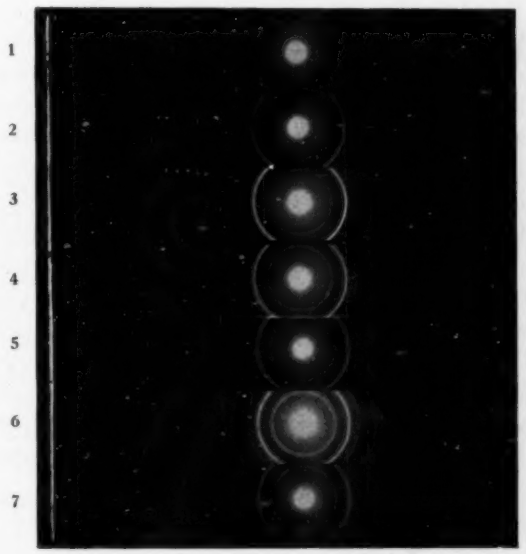


FIG. 6. Electron diffraction photographs of Bi of various thicknesses on Formvar substrates. 1: Formvar with about 17.8 atom layers of Bi; 2: Formvar with about 51.3 atom-layers of Bi; 3: Formvar with about 103 atom layers of Bi; 4: Formvar with about 172 atom layers of Bi; 5: Formvar with about 241 atom layers of Bi; 6: Formvar with about 309 atom layers of Bi; 7: Formvar with about 309 atom layers of Bi aged 20 hours.

proximating the points for the (202) ring. It is thought that the increase in diffuseness of the rings for the aging period of 1200 min. (shown by the dashed portions of the Fig. 5 curves) may be caused in part at least by condensation on the Bi films of gases and vapors in the apparatus.

The senior author (A. H. W.) acknowledges with gratitude assistance by a grant-in-aid from the Sigma Xi Research Fund.

Plastic Flow, Creep, and Stress Relaxation

Part IV. Anomalous Flow as an Order-Disorder Transition

CHARLES MACK
Technical and Research Department, Imperial Oil Limited, Sarnia, Ontario
(Received June 21, 1948)

Plastic systems show effects of steric hindrance at rest, which results in a molecule preventing its neighbors from occupying certain positions and introduces a certain degree of orientation. Under stress many more positions, resulting from the rotation, are forbidden to a molecule in motion. In order to make more positions available, the system must increase its volume under stress and changes from a state of greater order to a state of greater disorder. Based on this concept equations are given which present the strain-rate at constant temperatures as functions of stress, change in volume and degree of order and disorder. The changes in entropy and energy of interaction, accompanying the strained state, are expressed by a generalized partition function. This concept of anomalous flow is extended to visco-elastic effects for systems with rubber-like elasticity, and has been exemplified by data on polychloroprene under tension.

INTRODUCTION

ANOMALOUS flow in its widest meaning is connected with a non-linear relationship between shearing stress and shear-rate. In many cases the ratio of shearing stress to shear-rate, the so-called "apparent viscosity," decreases with increasing shearing stress, and several attempts have been made to explain this behavior Kauzmann and Eyring[1] suggested a breaking of chemical bonds in large molecules at high shear-rates. Such a process involves depolymerization and should therefore, not only be a function of the shearing stress but also of time as shown by Morris and Schnurmann.[2] These investigators using a special apparatus giving high shear-rates under laminar flow conditions found that long-chain molecules of M.W. above 100,000 in solution were broken down at shear-rates of 10^5 sec.$^{-1}$, and that the loss in viscosity increased with the number of cycles with which the experiment was repeated. Plastic systems, whose apparent viscosity decreases with increasing shearing stress and which are not thixotropic, do not show such a permanent loss in viscosity. Approaching the problem from a thermodynamic standpoint, Bondi[3] in an extensive study came to the conclusion that, at high shear-rates, corrections for a possible temperature-gradient are insufficient to explain the observed decrease in apparent viscosity. As an interpretation of these phenomena he proposed an orientation of the elements of flow parallel to the direction of flow. Bondi indicated that such an orientation should be accompanied by a decrease in the entropy change, and one would, therefore, rather expect an increase in viscosity in view of the change of the molecules from a state of disorder to a state of greater order. The writer[4] suggested a mechanism of anomalous flow based on a model consisting of elements of flow having different yield values. Although this theory explains the mechanical behavior of plastic substances, it does not take into account the possibility of self-diffusion which occurs even in

[1] W. Kauzmann and H. Eyring, J. Am. Chem. Soc. **62**, 3113 (1940).
[2] W. J. Morris and R. Schnurmann, Nature **160**, 674 (1947).
[3] A. Bondi, J. App. Phys. **16**, 539 (1943).
[4] C. Mack, J. App. Phys. **17**, 1086 (1946).

solid plastic materials such as metals, and in the following an attempt will be made to put this theory on a molecular basis.

ANOMALOUS FLOW AS AN ORDER-DISORDER REARRANGEMENT

Plastic systems are disperse systems and resemble solutions. In dilute solutions, the molecules of the solute are distributed at random and move freely in the solvent. At high concentrations, the molecules are not distributed at random, since each molecule occupies a definite volume, and statistics would require that certain elements of the total volume of the solution be occupied simultaneously by two molecules of the solute, which is materially impossible. Such solutions show effects of steric hindrance which results in a molecule preventing its neighbors from occupying certain positions.[5] Solutions of high concentrations are therefore not completely isotropic and exhibit orientation effects of the dissolved molecules. The degree of anisotropic distribution of orientation increases with increasing deviation from spherical shape of the dissolved molecule. Further anisotropic effects are introduced into the system by London-van der Waals' forces between the solute molecules and between the solute and solvent molecules and through the fact that the solvent possesses a certain degree of orderliness.

The theories of viscosity of liquids and solutions assume that the molecules are isolated, i.e., without action on each other, and that the molecules rotate in a field of laminar flow. This can occur only in dilute solutions below the limiting concentration at which isotropy exists. It has been shown above, that at higher concentrations certain positions are forbidden to the solute molecule at rest. If this assumption holds, it follows, that many more positions, resulting from the rotation, are forbidden to the solute molecule in motion. The end effect would then be a restriction of the motion, and in order to facilitate flow, more positions would have to be made available to the molecule in motion. Such an effect can be obtained only by an increase in the volume of the system, which has been reported by Scott Blair[6] on mixtures of bitumen and mineral powder and by Rehbinder[7] on metals. In the latter case microcracks were observed to develop under stress within the region of elastic deformations, which closed up again after release of the stress. An increase in volume of a body under stress is also indicated in all cases where Poisson's ratio, compressive strain/tensile strain, is smaller than 0.5, since, for a given tensile strain, the compressive strain decreases with decreasing value of Poisson's ratio.

According to Eyring's theory of viscosity,[8] holes exist in the liquid into which the molecules can jump. Such a jump occurs when the molecule has accumulated enough energy to surmount a potential barrier. The same concept can be applied to plastic systems, and the increase in volume under external forces provides for a sufficient increase in the volume of the holes to permit the molecules to jump into their new equilibrium positions. A molecule under consideration can jump across the potential barrier only, if, under the influence of its interaction with its neighbors, it requires a kinetic energy equal to or greater than the height of the barrier. After jumping across the barrier into a hole, the molecule would immediately return to its original position, if its kinetic energy should be fully restored. It must then be assumed that this additional energy is partly transmitted to the new neighbors, so that the molecule arrives at the hole with a kinetic energy smaller than that of the height of the barrier, which is insufficient for the return of the molecule to its original position. How many molecules have sufficient energy to pass by their neighbors depends upon the tightness of packing of the whole system, the force fields around their neighbors, the magnitude of the applied external force as well as temperature, and is proportional to the velocity gradient for the condition of shear.

Based upon the above considerations, a plastic system at rest consists of a number of molecules, the majority of which are confined to their positions under the influence of steric hindrance and cannot move to other positions, whereas the re-

[5] J. Duclaux, J. Polym. Science **II**, 626 (1947).

[6] Summary of the Proceedings of the Conference on the Rheology of Solids at Birmingham, Feb. 28, and March 1st, 1947, J. Scient. Indstr. **24**, 309 (1947).

[7] P. Rehbinder, Nature **159**, 866 (1947).

[8] H. Eyring and co-workers, J. Chem. Phys. **4**, 283 (1936); **5**, 726 (1937).

mainder is free to do so. Under stresses up to the yield value, the deformation is mainly elastic, i.e., a deformation during which the individual molecules retain their neighbors, and part of the energy introduced into the system is used to increase the volume of the holes. At stresses in excess of the yield value the molecule has accumulated enough energy to surmount the potential barrier and can move into a hole formed. The shift of a molecule from a confined position to a moving position represents a variation of the structure and reduces to a change of the degree of local order in the arrangement of the molecules.

For the order-disorder arrangement in plastic systems the model suggested by Lennard-Jones and Devonshire[9] may be applied, which consists of a molecule moving in the field of force of its neighbors and is confined by them to a small region of space referred to as cell. The essential difference between the ordered and disordered state is that in the former case the cell is localized, whereas in the latter it is not; in other words in the ordered or solid state the molecule cannot move from one cell to another, whereas in the disordered or liquid state it can. In the ordered state, the molecules will occupy, or rather vibrate about definite sites called α-sites. The disordered state will be referred to as β-sites, which are symmetrically related to the α-sites. The ordered state for plastic systems need not necessarily be a state of geometrical orientation as in crystals but may represent a state where the cell is localized. The degree of order of any configuration of the system can be described by the number of molecules N_α on the α-site and the number of molecules N_β on the β-site. The degree of order is defined by

$$n_\alpha = N_\alpha/N, \tag{1}$$

and of disorder by

$$n_\beta = N_\beta/N, \tag{2}$$

where N is the total number of molecules.

As a plastic system contains molecules of various sizes and a hole can accommodate only a fraction of a large molecule, a molecule or unit of flow will be referred to this fraction. For given values of N_α and N_β, the molecules are assumed to be distributed at random among the available sites, so that the states of order and disorder are uniform throughout the system, rendering the system as isotropic as a liquid from a macroscopic point of view.

The number of holes in a system is N_h, and if the volume of a hole is v_h, the number of holes per given volume is then

$$N_h = (V - V_s)v_h, \tag{3}$$

where V is the total volume of the system and V_s the volume of the molecules, such that the effective volume of a single molecule is V_s/N. Introducing this term into (3) leads to

$$N_h/N = (V - V_s)/V_s = \alpha, \tag{4}$$

giving the ratio of the number of holes to the number of molecules, which is assumed to be constant.

An increase in the volume of a hole is accompanied by the formation of a field force acting on the surface of the hole and tending to shift the surrounding molecules in a radial direction. The energy connected with this elastic strain is proportional to the external force applied. Since stresses up to the yield value cause elastic deformations during which the molecules retain their neighbors, the fields of forces surrounding the individual molecules and holes are in quilibrium, and a random distribution of the holes over α- and β-sites is conceivable. A hole on an α-site can then be considered as being localized at least for a certain time-interval. The number of holes on the α- and β-sites are then αN_α and αN_β.

At stresses in excess of the yield value, the process of anomalous flow, assuming the molecules to be rigid, consists in shifting a number of molecules on the α- as well as on the β-sites into the holes which are available on the β-sites. The postulate of holes on the α-sites permits of a possible shift of a molecule on the β-site into a hole on an α-site with subsequent closure of the cavity. As the number of holes represents only a fraction of the number of molecules, and a hole must be sufficiently large to accommodate a molecule, the number of molecules moving into the holes on the β-sites are relatively small at a given instance and stress. The product of the

[9] J. E. Lennard-Jones and A. F. Devonshire, Proc. Roy. Soc. **A169**, 317 (1939); **A170**, 404 (1939); Proc. Phys. Soc. (London) **52**, 729 (1940); also S. Glasstone, *Theoretical Chemistry* (D. van Nostrand Company, Inc., New York, 1944), p. 489.

number of jumps per unit time and of the distance traveled by one molecule represents the velocity of an individual molecule relative to that of its adjacent layers or its shear-rate. This rate is not necessarily the same for each molecule, but since the individual shear-rates as well as the energies involved in the jumps can be assumed to be additive, the average shear-rate, v_i, of a molecule will be equal to the total shear-rate divided by the number of molecules involved.

The stress σ applied to a plastic system under simple shear is the sum of two components, one causing flow and the other causing an increase in volume. With v the shear-rate and η the coefficient of viscosity, the viscous force necessary to shift the fraction of the molecules on the β-sites and part of the fraction of the molecules on the α-sites into the holes available at the β-sites is expressed by

$$\eta v_\beta = n_\beta \sigma - \alpha n_\beta K \Delta V / V_0, \tag{5}$$

where αn_β = fraction of holes on the β-sites, K = bulk modulus, ΔV and V_0 = change in volume and volume of the system at rest.

As a molecule moves into a hole on the β-site, it leaves a hole behind; the molecules move therefore in one direction and the holes in the opposite direction. A possible shift of a fraction of the molecules into the holes on the α-sites is represented by

$$-\eta v_\alpha = -n_\alpha \sigma + \alpha n_\alpha K \Delta V / V_0. \tag{6}$$

The sum of both terms gives

$$\eta(v_\beta - v_\alpha) = \eta v = (n_\beta - n_\alpha)\sigma - \alpha(n_\beta - n_\alpha) K \Delta V / V_0, \tag{7}$$

where v is the observed shear-rate. The force $K\Delta V/V_0$, necessary for the formation of holes, represents a hydrostatic tension and is according to the theory of elasticity equal to $3G\gamma$ where G = modulus of rigidity, γ = shear-strain, and Eq. (7) can be written

$$\eta v = (n_\beta - n_\alpha)(\sigma - 3\alpha G\gamma). \tag{8}$$

Dividing by η and replacing G/η by τ the relaxation time, changes this equation into:

$$v = (n_\beta - n_\alpha)(\sigma/\eta - 3\alpha\gamma/\tau). \tag{9}$$

The first term on the right side represents the sum of the individual viscous shear-rates, and the other expresses the sum of the rates at which the individual holes are formed or closed and shifted to a new position. Each hole must remain stationary for a certain time during which a molecule gathers enough energy to shift into the hole. The relaxation time, therefore, acquires a novel meaning, in that it becomes the sum of the average time of residence of a hole and the time required to shift it.

Replacing $G\gamma$ by σ and 3α by $\beta = 1-b$, gives the following expression for the coefficient of viscosity of anamolous flow:

$$\eta v = (n_\beta - n_\alpha)(1-\beta)\sigma = (n_\beta - n_\alpha) b\sigma. \tag{10}$$

This equation is essentially the same as that previously given,[4] however the above treatment describes anomalous flow at constant temperature as a change in the order-disorder arrangement of the molecules. Such a change must be accompanied by a change in the field of force around a molecule and a change in entropy, which is comparable to the entropy of fusion, if the number of molecules on the β-sites increases, and comparable to the entropy of crystallization, if the number of molecules on the α-sites increases.

PARTITION FUNCTION FOR ORDER-DISORDER ARRANGEMENT

The number of neighbors of a molecule on any site is z, which is 6 for cubical packing and 12 for close packing. The number of β-sites occupied by the molecules and holes adjacent to any molecule on an α-site is $(1+\alpha)zn_\beta$; similarly the number of occupied α-sites around any occupied β-site will be $(1+\alpha)n_\alpha z$. The energy of interaction of each pair of molecules on adjacent α- and β-sites will be designed by e. This energy, being the resultant of the repulsive and attractive forces, is a function of the distance between the two sites and should, therefore, be a function of the volume as a whole. The attractive and repulsive forces, however, are also functions of the degree of order or disorder, and the energy of interaction will also depend upon this degree. The average energy e_α of a molecule on an α-site as a result of its $(1+\alpha)n_\beta z$ neighbors on β-sites will thus be given by

$$e_\alpha = (1+\alpha)n_\beta z e, \tag{11}$$

and the average energy e_β of a molecule on a

β-site by

$$e_\beta=(1+\alpha)n_\alpha ze. \tag{12}$$

The average energy required to remove a molecule from an α-site and transfer it to a β-site is

$$\Delta e=e_\beta-e\alpha=(1+\alpha)(n_\alpha-n_\beta)ze. \tag{13}$$

The total energy of interaction E is given either by $(1+\alpha)N_\alpha e_\alpha$ or $(1+\alpha)N_\beta e_\beta$, which must be identical, hence

$$E=(1+\alpha)N_\beta e_\beta=(1+\alpha)N_\alpha e_\alpha \\ =(1+\alpha)^2Nzn_\alpha n_\beta e. \tag{14}$$

The number of ways of distributing the molecules and holes among the sites so that there are N_α molecules and αN_α holes on the α-sites and N_β molecules and αN_β holes on the β-sites is

$$\Omega=\frac{(N+\alpha N)!}{(N-N_\alpha)!N_\alpha![\alpha(N-N_\alpha)]!(\alpha N_\alpha)!} \\ \cdot\frac{(N+\alpha N)!}{(N-N_\beta)!N_\beta![\alpha(N-N_\beta)]!(\alpha N_\beta)!} \\ =\left\{\frac{[(1+\alpha)N]!}{N_\alpha!N_\beta!(\alpha N_\alpha)!(\alpha N_\beta)!}\right\}^2. \tag{15}$$

Combining Eqs. (14) and (15) leads to the following partition function

$$Q=Q''Q'=\Omega\exp(-E/kT)Q', \tag{16}$$

where Q'' and Q' represent the partition functions of the system in the disordered and ordered state. Making use of the thermodynamic relation, $A=E-TS$, where A = maximum work function, E = total energy and S = entropy, the partition function may be written

$$A=-kT\ln\Omega+E-kT\ln Q'.$$

Introducing the values for Ω and E from Eqs. (14) and (15) and applying Stirling's approximation for the logarithm of a factorial, $\ln N!=N\ln N-N$, gives

$$A=2kTN[(1+\alpha)n_\alpha\ln n\alpha+(1+\alpha)n_\beta\ln n_\beta \\ -(1+\alpha)\ln(1+\alpha)+\alpha\ln\alpha] \\ +(1+\alpha)^2Nzn_\alpha n_\beta e-kT\ln Q'. \tag{17}$$

The first term on the right side represents the product of the temperature and the entropy of random mixing of the molecules and holes on the α- and β-sites, which is comparable to the entropy of fusion. For anomalous flow at constant temperature, the change in the energies is of importance only, hence

$$\Delta A-\Delta A_0=2kTN(1+\alpha)[n_\alpha\ln n_\alpha+n_\beta\ln n_\beta \\ -n_{\alpha0}\ln n_{\alpha0}-n_{\beta0}\ln n_{\beta0}]+(1+\alpha)^2Nzen_\alpha n_\beta \\ -(1+\alpha)^2Nze_0n_{\alpha0}n_{\beta0}, \tag{18}$$

where the subscripts 0 refer to the conditions at the yield value.

At constant temperature the work function consists of the work done in shifting the molecules and the work done for the formation of the holes. The work done in shifting a molecule is $f\lambda$, where f if the force applied and λ the distance jumped. Multiplying and dividing by a, the area of a molecule, leads to the following expression for the work done in shifting N_β molecules to the β-sites and N_α molecules to the α-sites.

$$\Delta W=(N_\beta-N_\alpha)fa\lambda/a.$$

The term f/a is equal to the shearing stress σ and $a\lambda$ corresponds to the volume of a molecule, hence $Na\lambda$ represents the unit volume containing N molecules, and

$$\Delta W=(n_\beta-n_\alpha)\sigma. \tag{19}$$

The force connected with the hole formation is according to Eqs. (8) and (10) equal to $3\sigma=(1-b)\sigma/\alpha$, and for one hole equal to $(1-b)\sigma/\alpha^2N$. The work done in forming αN_β holes on the β-sites and closing αN_α holes on the α-sites is

$$p\Delta V=(N_\beta-N_\alpha)(1-b)\sigma\Delta V/\alpha N.$$

But $\Delta V/\alpha N$ is the volume of a hole and must be large enough to accommodate a molecule, hence this volume is of the same order as a λ, and

$$p\Delta V=(n_\beta-n_\alpha)(1-b)\sigma. \tag{20}$$

Combining Eqs. (19) and (20) gives for the work function

$$\Delta A-\Delta A_0=\Delta W-\Delta W_0-p\Delta V+p_0\Delta V \\ =(n_\beta-n_\alpha)b\sigma-(n_{\beta0}-n_{\alpha0})b\sigma_0, \tag{21}$$

and the work functions become numerically equal to the effective stresses.

Equation (18) may now be written as follows:

$$\Delta A-\Delta A_0=-T(\Delta S-\Delta S_0)+E-E_0,$$
$$(n_\beta-n_\alpha)b\sigma-(n_{\beta 0}-n_{\alpha 0})b\sigma_0=2kTN(1+\alpha)$$
$$\times(n_{\alpha 0}\ln n_{\alpha 0}+n_{\beta 0}\ln n_{\beta 0}-n_\alpha\ln n_\alpha-n_\beta\ln n_\beta)$$
$$+N(1+\alpha)^2zn_\alpha n_\beta e-N(1+\alpha)^2zn_{\alpha 0}n_{\beta 0}e_0. \quad (22)$$

Differentiation with respect to n_β gives for the condition of constant temperature and stress

$$2b\sigma=-T\partial\Delta S/\partial n_\beta+N(1+\alpha)^2ze(n_\alpha-n_\beta).$$

At $\sigma=\sigma_0$, the yield value, the entropy term is zero, hence

$$2b\sigma_0=N(1+\alpha)^2ze_0(n_{\alpha 0}-n_{\beta 0}),$$

which yields immediately an expression for the energy of interaction in terms of the yield value, pending upon whether $n_{\alpha 0}$ is larger or smaller than $n_{\beta 0}$,

$$N(1+\alpha)^2ze_0=\pm 2b\sigma_0/(n_{\beta 0}-n_{\alpha 0}). \quad (23)$$

Making use of Eq. (17) and taking only the disordered state into consideration leads to an expression for the entropy of random mixing in terms of the yield value. For this purpose, there must be added to the work done in shifting the molecules and forming the holes the contribution U_0 resulting from the volume increase of the system as a function of the elastic deformation up to the yield value, hence

$$(n_{\beta 0}-n_{\alpha 0})b\sigma_0-U_0=2kTN(1+\alpha)[n_{\alpha 0}\ln n_{\alpha 0}$$
$$+n_{\beta 0}\ln n_{\beta 0}-\ln(1+\alpha)+(\alpha/1+\alpha)\ln\alpha]$$
$$+2b\sigma_0n_{\alpha 0}n_{\beta 0}/(n_{\beta 0}-n_{\alpha 0}). \quad (24)$$

Keeping in mind that

$$n_\alpha n_\beta=n_\beta(1-2n_\beta)=[1-(2n_\beta-1)^2]/4$$

gives the following term for the entropy of random mixing at the yield point:

$$T\Delta S_0=-2kTN(1+\alpha)[n_{\alpha 0}\ln n_{\alpha 0}+n_{\beta 0}\ln n_{\beta 0}$$
$$-\ln(1+\alpha)+(\alpha/1+\alpha)\ln\alpha]=0.5b\sigma_0$$
$$\times[3(n_{\alpha 0}-n_{\beta 0})+1/(n_{\beta 0}-n_{\alpha 0})]+U_0. \quad (24a)$$

In this expression all terms can be determined experimentally except $n_{\beta 0}-n_{\alpha 0}$ and it is of necessity to find such a relationship. It has been mentioned above that the energy of interaction e is a function of the degree of disorder. If two different degrees of disorder are not too far apart, it can be assumed that the variation in the energies of interaction is so small as to be negligible. Such a condition exists, if the degrees of disorder, $n_{\beta 0}$ in one case and $n_\beta=\frac{1}{2}$ in the other are close together, a condition which is usually met at stresses near the yield value. Combining Eqs. (22) and (24) for this condition yields

$$(n_\beta-n_\alpha)b\sigma=0=-2kTN(1+\alpha)$$
$$\times[\ln(2+2\alpha)-(\alpha/1+\alpha)\ln\alpha]$$
$$+b\sigma_0/2(n_{\beta 0}-n_{\alpha 0})+U_0n_{\beta 0}-n_{\alpha 0}$$
$$=b\sigma_0/\{4kTN(1+\alpha)$$
$$\times[\ln(2+2\alpha)-(\alpha/1+\alpha)\ln\alpha]-2U_0\}.$$

The term U_0 is small in comparison to the remainder in the brackets, and since it also compensates for the assumed small variation in the energies of interaction, it can be omitted without introducing a serious error. Replacing N, the number of molecules per cm³, by $\rho Na/\mathrm{M}$, where ρ=density, Na=Avogadro's number and M=molecular weight, gives

$$n_{\beta 0}-n_{\alpha 0}\approx b\sigma_0\mathrm{M}/4RT\rho(1+\alpha)$$
$$\times[\ln(2+2\alpha)-(\alpha/1+\alpha)\ln\alpha]. \quad (25)$$

With the knowledge of the value of $n_{\beta 0}-n_{\alpha 0}$ and the value of $n_\beta-n_\alpha$ obtainable from the relationships between stress and shear-rate, the various energy terms in Eq. (22) can be calculated. If the molecular weight is not known, Eq. (23) may be used for an estimation of $n_{\beta 0}-n_{\alpha 0}$

$$n_{\beta 0}-n_{\alpha 0}=2b\sigma_0/N(1+\alpha)^2ze_0.$$

At the yield value, it can be assumed that $N(1+\alpha)e_0$, representing the energy of interaction between $N(1+\alpha)$ pairs of molecules and holes is equal to $2\sigma_0$, hence

$$n_{\beta 0}-n_{\alpha 0}=b/z(1+\alpha). \quad (26)$$

The value of $n_{\beta 0}-n_{\alpha 0}$ thus is a function of z, the degree of packing, which in many cases can be estimated.

PLASTIC FLOW

Plastic flow, as previously defined,[4] is a deformation mechanism having a curvilinear relationship between shearing stress and shear-rate and a constant rate of shear at constant stress. For this type of anomalous flow the following relationship has been suggested by de Waele,[10]

[10] A. de Waele, J. Oil and Coll. Chem. Assoc. 4, 33 (1923).

Farrow,[11] and Wo. Ostwald:[12]

$$\sigma/\sigma_0=(v/v_0)^b, \tag{27}$$

which can be used directly in connection with Eq. (10),

$$\eta v=(n_\beta-n_\alpha)b\sigma$$

to give the following expression for $n_\beta-n_\alpha$ as a function of the stress,

$$(n_\beta-n_\alpha)/(n_{\beta 0}-n_{\alpha 0})=(\sigma/\sigma_0)^{(1-b)/b}. \tag{28}$$

The value of $n_\beta-n_\alpha$ and, therefore, the degree of disorder increases with increasing shearing stress. In cases where Eq. (27) applies, the change in entropy should increase with increasing stress, and the deformation should be accompanied by an absorption of heat. Rearranging Eq. (22) gives for this case the following expression for the entropy change

$$\begin{aligned}T(\Delta S-\Delta S_0)&=2kTN(1+\alpha)[n_{\alpha 0}\ln n_{\alpha 0}\\&+n_{\beta 0}\ln n_{\beta 0}-n_\alpha\ln n_\alpha-n_\beta\ln n_\beta]\\&=(n_\beta-n_\alpha)b\sigma+N(1+\alpha)^2ze n_\beta n_\alpha\\&-0.5b\sigma_0[3(n_{\alpha 0}-n_{\beta 0})+1/(n_{\beta 0}-n_{\alpha 0})].\end{aligned} \tag{29}$$

Another equation which covers a curvilinear relationship between shearing stress and shear-rate has been given by Ludwik[13] and Prandtl[14] and may be expressed by

$$\exp[(\sigma-\sigma_0)/\sigma_0]=(v/v_0)^b, \tag{30}$$

which in connection with Eq. (10) leads to

$$\begin{aligned}(n_\beta-n_\alpha)/(n_{\beta 0}-n_{\alpha 0})\\=(\sigma_0/\sigma)\exp[(\sigma-\sigma_0)/b\sigma_0].\end{aligned} \tag{31}$$

In this case the degree of disorder also increases with increasing stress.

STRESS-HARDENING

In the previous section the cases of an increase in the degree of disorder with increasing stress have been discussed. With values of b in Eq. (27) larger than 1, there is a disproportionally small increase in the shear-rate with increasing shearing stress. Since the shear-rate is proportional to the number of molecules shifting into new positions, this number must decrease with increasing shearing stress. The same holds even more so for cases where b is negative (dilatancy), as the shear-rate decreases with increasing shearing stress. In both cases a deformation increases the number of molecules on the α-sites, and with increasing degree of order, the strength of the system is increased, i.e., the stress that can be supported, hence the term stress-hardening given to the mechanical behavior of such systems.

Stresses up to the yield value cause an increase in the volume of the holes, into which the molecules can slip. With the number of shifting molecules decreasing with increasing stresses, it can be assumed, that the molecules in the ordered state remain on the α-sites, and that the molecules on the β-sites slip into the holes available on the α-sites and become there localized. Such a process is connected with a decrease in the volume of the holes, and Eq. (8) may be written

$$\eta v=n_\beta(\sigma+3\alpha G\gamma),$$

or with

$$\begin{aligned}1+3\alpha&=b,\\\eta v&=n_\beta b\sigma.\end{aligned} \tag{32}$$

Connecting with Eq. (27) gives n_β as a function of the stress

$$n_\beta/n_{\beta 0}=(\sigma_0/\sigma)^{(b-1)/b}. \tag{33}$$

For the case of b being negative, the shear-rate decreases with increasing stress, and dv becomes negative. Since η is a function of $d\sigma/dv$, Eq. (32) may be written

$$-\eta v=-n_\beta b\sigma,$$

resulting in the following expression for n

$$n_\beta/n_{\beta 0}=(\sigma_0/\sigma)^{(1+b)/b}. \tag{34}$$

With increasing degree of order, the change in entropy is negative in both cases.

WORK-HARDENING

An increase in strength of the system can also be achieved through work-hardening. In this case the shearing stress becomes a function of the shear-rate and the time (time-hardening) or a function of the shear-rate and shear-strain (strain-hardening). Among the various equations, which have been found to be applicable to experi-

[11] F. D. Farrow and G. M. Lowe, J. Text. Inst. **14T**, 414 (1923).
[12] Wo. Ostwald, Kolloid Zeits. **36**, 99 (1925).
[13] P. Ludwik, *Elements der Technologischen Mechanik* (Verlagsbuchhandlung, Julius Springer, Berlin, 1909).
[14] E. Prandtl, Zeits. f. augew. Math. u Mech. **8**, 85 (1928).

mental data,[15] the following may be presented as typical examples for time- and strain-hardening:

$$\sigma/\sigma_0=(vt/v_0t_0)^b, \tag{35}$$

and

$$\sigma/\sigma_0=(v\gamma/v_0\gamma_0)^b, \tag{36}$$

where t and γ are the time and shear-strain. In these equations the products of shear-rate and time or shear-rate and shear-strain are constant at constant shearing stress, but the shear-rate decreases with time or shear-strain. In both cases the number of molecules on the β-sites decreases with increasing time or shear-strain in a similar manner as discussed for stress-hardening, and

$$n_\beta/n_{\beta 0}=(\sigma/\sigma_0)^{(1-b)/b}t_0/t, \tag{37}$$

$$n_\beta/n_{\beta 0}=(\sigma/\sigma_0)^{(1-b)/b}\gamma_0/\gamma, \tag{38}$$

where b may be also larger than 1 or negative.

STRENGTH OF PLASTIC SYSTEMS

Plastic systems whose shear-rate increases proportionally more than the shearing stress undergo a change from a state of lesser disorder to a state of greater disorder, and their strength is, therefore, determined by the yield value. Plastic systems, subject to hardening under the influence of stress, time or strain on the other hand, shift to a state of greater order with increasing stresses, and reach maximum orientation when the number of molecules on the β-sites become a minimum. At this point the system behaves like a rigid body at least in the direction of stress, and the yield value has increased. Any further increase in stress will tend to increase the volume of the holes and thus allow the system to return from a state of order to a state of disorder resulting finally in complete failure. The maximum strength is therefore obtained at the maximum yield value, the conditions for which have been discussed in Part II of this series[15] in terms of stress and strain.

VISCO-ELASTIC EFFECTS

Systems showing visco-elastic effects are mainly of two types. In the first case the system consists of a rigid network enmeshing a liquid phase, and a stress causes an elastic deformation of the network and a viscous or plastic deformation of the liquid phase. If the molecules of the network retain their neighbors during the deformation, the latter is small and purely elastic, and such systems have been discussed earlier.[16] The other type, usually referred to as rubber-elasticity, represents a system consisting of flexible chain molecules. The segments of these large molecules are assumed to be distributed at random over the α- and β-sites. At shearing stresses of small magnitude, the deformation is purely elastic and obeys Hooke's law. At greater stresses the segments on the β-sites perform rotational and translatory motions and are shifted into the α-sites. The deformation connected with this transition from one site to another consists of two parts. One part, being translatory, is a function of time and is therefore a viscous deformation; the other part consists of an elastic deformation, which the segment undergoes after its arrival at the α-site. The viscous deformation is accompanied by the formation of holes, which are closed again after the molecules have occupied the α-sites. The net result is a constancy of the volume which has been observed for such systems under stress.

The shearing stress applied to such a system consists of two components, one connected with the viscous deformation and the other with the elastic deformation, and $\sigma=\sigma_\beta+\sigma_\alpha$, where the suffixes refer to the α- and β-sites. At constant stress, immediately after its application, the stress is carried by the molecules on the β-sites, and $\sigma=\sigma_\beta$. With the shift into the α-sites, a part of the stress is taken over by the molecules on the α-sites, and σ_β decreases from σ to 0, at which point $\sigma=\sigma_\alpha$. Making use of Eq. (5), the stresses

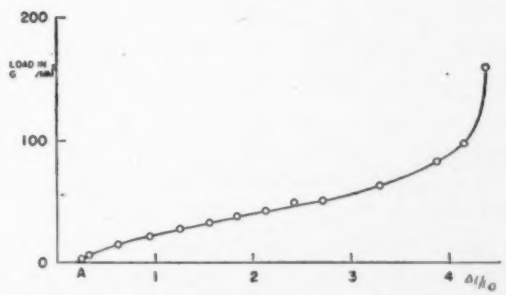


FIG. 1. Load/original area as a function of elongation for polychloroprene. Reproduced from J. Phys. and Coll. Chem. 51, 70 (1947).

[15] C. Mack, J. App. Phys. 17, 1093 (1946).

[16] C. Mack, J. App. Phys. 17, 1101 (1946).

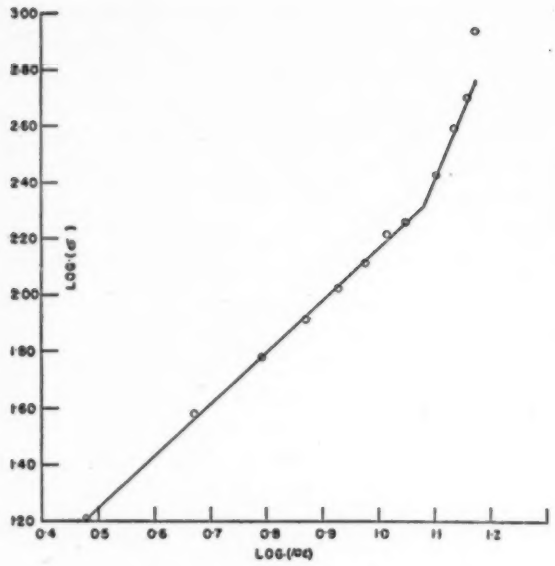


FIG. 2. Log stress−log strain relationship of polychloroprene.

involved with the viscous and elastic deformation for molecules shifting are

$$n_\beta\sigma_\beta=\eta v+\alpha n_\beta K_\beta(\Delta V)/V_0=\eta v+n_\beta\beta\sigma_\beta,$$
$$n_\beta\sigma_\alpha=G\gamma-\alpha n_\beta K_\alpha(\Delta V)/V_0=G\gamma-n_\beta\beta\sigma_\alpha.$$

Their sum on rearranging is

$$G\gamma+\eta v=n_\beta(\sigma_\alpha+\sigma_\beta)+n_\beta\beta(\sigma_\alpha-\sigma_\beta),$$

or since $\sigma_\alpha-\sigma_\beta=\sigma-2\sigma_\beta$, substituting b for $1+\beta$

$$G\gamma+\eta v=n_\beta b\sigma-2n_\beta\beta\sigma_\beta. \tag{39}$$

The elastic strain γ and the strain-rate v can be power functions or exponential functions of the corresponding stresses as in the case of plastic flow, and the experiments have to be properly chosen for the evaluation of these functions. If the experiment is carried out by keeping the applied stress constant until equilibrium is reached, v and σ_β become zero and

$$G\gamma=n_\beta b\sigma. \tag{40}$$

On unloading the system returns to an unstrained state and original degree of disorder, and for this condition Eq. (39) also holds with the signs reversed. In view of the high molecular weight of such systems, the viscosity term has a large value, and the return to the original state of disorder is a slow process. For a rapid return, as in the case of rubber, another term $\eta'v'$ must be added with $\eta'\ll\eta$ and $v'\gg v$ in order to counteract the internal friction introduced by the high viscosity. In this case we are confronted with a more complex structure consisting of fractions of molecules, which become oriented on stretching, and fractions of other molecules, which remain in the disordered state. This concept of a two-phase system is borne out by Hauser's morphological investigations by means of the ultramicroscope and electrone-microscope[17] and is in agreement with Field's x-ray studies of stretched rubber.[18] In the latter case the amount of crystalline rubber was found to be 80–85 percent at elongations of approximately 600 percent and to decrease with elongations of greater magnitude in agreement with the above discussion on the strength of materials. For such systems Eq. (39) becomes

$$G\gamma+\eta v+\eta'v'=n_\beta b\sigma-2n_\beta\beta\sigma_\beta+n_\beta'\sigma, \tag{41}$$

where n_β' refers to the fraction of molecules which remain in a state of disorder.

As an illustration of the mechanism of rubber-like elasticity in particular and of the anomalous deformation as an order-disorder rearrangement in general, reference may be made to the data obtained by Kinell[19] on the deformation of polychloroprene under tension. These data were chosen, because, for loading as well as unloading, the load was kept constant for a sufficient time until equilibrium conditions were obtained. The load per original area as a function of the elongation for the condition of unloading at 29.2°C is presented in Fig. 1 for Kinell's sample III, which consisted of unvulcanized polychloroprene obtained from emulsion polymerization in presence of 2 percent of sulfur. The distance OA corresponds to the non-recoverable deformation. Since the deformations are large, the natural strains, $(\ln l/l_0)$, must be substituted for the conventional strains, $(\Delta l/l_0)$, and they have to be further corrected for the non-recoverable deformation. The load per original area can be transformed into the actual stress by multiplication with (l/l_0), since the change in volume is practically nil. With these corrections the first three points in

[17] E. A. Hauser and co-workers, Ind. Eng. Chem. **36**, 634 (1944); J. Phys. and Coll. Chem. **51**, 278 (1947).
[18] J. E. Field, J. App. Phys. **12**, 23 (1941).
[19] Per-Olof Kinell, J. Phys. and Coll. Chem. **51**, 70 (1947).

Fig. 1 give a linear relationship between stress and strain with a modulus of 5.25×10^6 dynes/cm^2. From the third point on the relationship between stress and strain is

$$\sigma/\sigma_0=(\epsilon/\epsilon_0)^b, \tag{42}$$

where σ refers to the tensile stress, and $\epsilon=(\ln l/l_0)$ is the natural tensile strain. This relationship is represented as a plot of $\log\sigma$ *versus* $\log\epsilon$ in Fig. 2. It will be observed that this relationship has two values of b, being 1.84 for the first part and 4.68 for the upper part. The absence of continuity may be explained by the circumstance, that at the point of inflection the polychloroprene is in a highly ordered state, and the number of available holes is small. In order to shift the remainder of the molecules on the β-sites into the α-sites, a hole would have to be available for each molecule. In view of the restricting influence of its neighbors, a molecule on a β-site would require a somewhat larger volume in order to carry out rotational motion. Thus a large part of the energy is required to increase the volume necessary for the shift, and the number of holes per molecule or their volume, $\alpha=(b-1)/3$, increases in this case from 0.28 to 1.23. This is also borne out by the increase in the "modulus of rubber-elasticity" from 5.31×10^6 to 14.02×10^6 dynes/cm^2 for the upper part.

The individual segment of rubber was reported by Treloar[20] to have a M.W. of 4100 corresponding to 60 isoprene units. Assuming the same number of units in a segment of polychloroprene, the M.W. will be 5310. The value of $n_{\beta_0}-n_{\alpha_0}$ for the point, where the rubber-elasticity commences to manifest itself, is according to Eq. (25) equal to 0.09933, and since $n_{\beta_0}-n_{\alpha_0}=2n_{\beta_0}-1$, 54.97 percent of the molecules have occupied β-sites and 45.03 percent are on the α-sites. At the point of inflection corresponding to a stress of 22.9 kg/cm^2 16.4 percent of the molecules occupy β-sites, and at the last point of the upper straight line 8.8 percent of the molecules are on the β-sites. This point appears to be the upper limit, to which this sample can be stretched, and a further increase in stress will throw the system in an increasingly disordered state. With the knowledge of the values of n_β, the following data are obtained from Eq. (24): $T\Delta S_0=1.466\times10^7$ ergs. $U_0=3.77\times10^5$ ergs, $T\Delta S$ at the point of inflection $=1.17\times10^7$ ergs, hence $T(\Delta S-\Delta S_0)=-2.92\times10^6$ ergs. The process of stretching is therefore accompanied by a temperature rise, as it should be. For the condition of unloading, the entropy change corresponding to this value is 1.23 cal. per degree per mole, which is of the order of magnitude of the entropy of fusion for substances which rotate both in the liquid and solid state.[9]

The above theory represents only a general outline of anomalous flow as an order-disorder transition, and is primarily based on measurable quantities such as stress and strain-rate. When considering definite plastic systems further refinements can be introduced, and the partition function would have to be correspondingly extended.

Note added in proof: Through an error the force $K\Delta V/V_0$ was put equal to $3G\gamma$. The sentence following Eq. (7) should be amended as follows: The force $K\Delta V/V_0$ necessary for the formation of the holes, represents a hydrostatic tension and is according to the theory of elasticity equal to $G\gamma$ etc. The subsequent terms of $b=1\pm3\alpha$ should be replaced by $b=1\pm\alpha$. With this change the data for the polychloroprene become as follows: $n_{\beta_0}-n_{\alpha_0}=0.03418$; n_β at the point of inflection 0.1543 and at the last point 0.0831; $T\Delta S_0=4.275\times10^7$; $U_0=1.31\times10^5$; $T\Delta S$ at the point of inflection $=3.55\times10^7$; $T(\Delta S-\Delta S_0)=-7.247\times10^6$ ergs.

[20] L. R. G. Treloar, Trans. Faraday Soc. **40**, 59 (1944).

Letters to the Editor

Note on "Statistical Aspects of Fracture Problems"

[Benjamin Epstein, J. App. Phys. 19, 140 (1948)]

FRANKLIN H. FOWLER, JR.
Lessells and Associates, 916 Commonwealth Avenue, Boston, Massachusetts

June 16, 1948

TO supplement the interesting points on "Statistical aspects of fracture problems" by Benjamin Epstein, the writer would like to mention three additional topics which are also important in this field.

First, some theoretical considerations of the nature of the statistical distribution of specimen strengths, based on the assumption that strength variations result from variations in microstructure, can be made. There would first be the strength of a specimen with optimum microstructure. The probability of a stronger specimen would then be zero. Second, the weakest possible specimen is one which would fall apart at zero stress. Therefore, there is a lower limit equal to or greater than zero to the range of possible strengths. For these reasons it would seem desirable to give more consideration to distribution functions, which, like the Pearson type I curve, drop continuously to zero probability at the finite limits to the range of possible stress.

A further consideration of the nature of the probability distribution curve leads to the analysis of an effective inspection process. When an effective inspection process exists, there is a minimum passable strength higher than the minimum strength without an inspection process. In this case, the probability distribution curve for a particular specimen will jump from zero to a definite value at the minimum passable strength, and the remainder of the new probability distribution function will be proportional to the original one.

These considerations are of interest in an appraisal of the influence of size effect on design. First, there is a definite limit to the gain in fracture stress obtainable by reducing the size of the specimen. Second, there is a limit to the natural statistical reduction in fracture stress resulting from an increase in size. This is the minimum possible fracture stress of a specimen of any size. The minimum possible fracture stress can be increased when an effective inspection process can be introduced, but the number of rejects will increase as the specimen size increases.

The second topic is an additional "bundle of threads" problem discussed by Weibull.[1] In this problem the extension rather than the load is fixed. The strength of the bundle is then fixed by the strength of the strongest strand, and we have another problem in the distribution of extreme values.

The third topic is the desirability of basing statistical strength on the stress level above which a given percentage of specimens would fail rather than basing it on the mode. The designer should concern himself with the probability (zero if he can get it) that his structure will fail rather than the number of other structures which will fail under similar conditions of loading. Also, for strength of material tests of different types, it is sometimes reasonable to assume that the specimens in the nth percentile of the group subjected to one test are equivalent to those in the nth percentile of the group subjected to a different test. When this can be done, the difficulties associated with the impossibility of twice testing a given specimen to destruction are compensated for. For these reasons, it seems desirable to favor the use of percentiles as a basis for statistical strength in preference to the use of such other statistical characteristics as the mode or as the expected strength.

[1] Weibull, "The phenomenon of rupture in solids," Ing. Vetenskaps Akad. Handl. No. 153, p. 40.

Transmission Properties of Materials in the Millimeter Wave Region*

E. WANTUCH, J. H. ROHRBAUGH, J. P. COOLEY, J. H. GREIG, AND J. SHMOYS
New York University, New York, New York

July 27, 1948

THE transmission properties of polystyrene sheets have been investigated, using a grating spectrometer. The sources developed in our laboratory allow measurements over the wave-length range from 0.3 mm to 2.2 mm. Curve *A* on Fig. 1 shows the percent transmission of a $\frac{1}{4}$-inch sheet plotted *versus* wave-length. Curve *B* shows the percent transmission of two $\frac{1}{4}$-inch sheets placed next to each other. These two samples were cut from the same sheet which had been purchased from the Amphenol Corporation. Curve *C* shows the transmission characteristic of a $\frac{3}{8}$-inch sheet of polystyrene purchased from the Plax Corporation. The position of the edge of the absorption band of curve *C* is slightly shifted compared to the other two curves, and the percent transmission on the low wave-length side is materially different. This behavior may be accounted for by different processes of manufacture or content of impurities.

The transmission of $\frac{1}{4}$-inch thick Teflon, supplied by du Pont, is shown in Fig. 2. The resolving power was

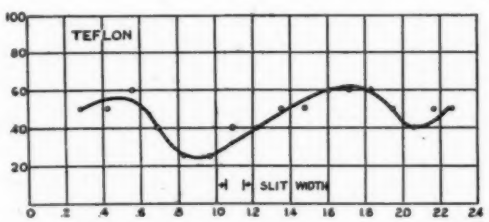


FIG. 1. Percent transmission of Teflon in the wave-length range from 0.3 to 2.2 cm.

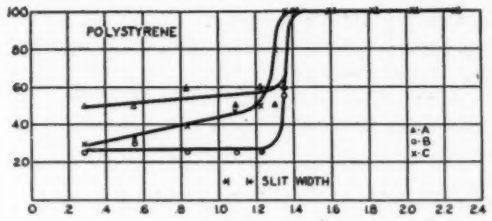


FIG. 2. Percent transmission of polystyrene in the wave-length range from 0.3 to 2.2 cm.

limited to a value of 10 by slit width. The accuracy of all transmission data is estimated to be 10 percent transmission. A more detailed discussion of these experiments will be published later.

* This work was supported by the Air Materiel Command, Watson Laboratories, under Contract No. W28-099 ac-171.

Some Remarks Concerning a Proposed Method for Testing the Brittleness of Elastomers

C. H. KLUTE
Shell Development Company, Emeryville, California
August 5, 1948

THE quality of brittleness being of prime importance where elastomers are concerned, it is not surprising that considerable space in the literature has been devoted to the analysis of the concept of brittleness and to empirical methods of testing for it. The consensus is that brittle fracture may be realized in a normally ductile elastomer provided the rates of deformation are sufficiently high. In particular there exists both experimental and theoretical evidence to the effect that for elastomers at a given temperature, a critical velocity of deformation can be reached beyond which brittle fracture will always occur. For details of this aspect of the problem the reader is referred to an excellent review which appeared recently.[1] Accordingly, a number of brittleness testing procedures have been devised in which a specimen is bent either around a mandrel or according to its own elastic curve.[2] In both cases the stretched side of the specimen is extended at a rate depending upon the rate of bending. When the rate of extension becomes sufficiently great, the specimen will presumably fail in tension. One notes that bending a specimen about a curve of specified radius is a convenient method for obtaining a reproducible rate of extension for that part of the specimen which is on the outside of the curve and also that the maximum extension achieved can be made to remain well below the ultimate elongation of the material at low rates. Although known strain rates might be applied to a specimen of elastomer by means of a cam-operated stretching device, the bending method herein described offers some experimental advantages, particularly in the avoidance of clamping stresses.

In what follows, the conditions of strain imposed upon a thin elastic strip when subjected to a hypothetical brittle-point test are given approximately. The proposed brittle-point test would impose mathematically describable strain rates upon specimens of elastomers at a series of low temperatures until a temperature is reached at which the imposed strain rate causes the fracture of the specimen. The strain rates are to arise as a result of the bending that a strip of elastomer would undergo if it were caused to travel lengthwise along a curvilinear path with known velocity. To derive these conditions it is necessary to accept some of the usual assumptions required to derive the simple flexural formula and, in addition, to assume that near the brittle point the specimen will be substantially unable to transmit tensile strains from one segment to the next rapidly, compared to the rate at which the tensile strains are applied to the segment. In such a system the strain rate would depend upon the position of the segment along the path selected, as well as the velocity along the path. If ϵ were the tensile strain experienced by the extended side of a segment of an elastic strip of thickness θ at a distance s along the path at which the radius of curvature was ρ, the time derivative of strain may easily be shown to be:

$$\frac{d\epsilon}{dt}=\frac{d\epsilon}{ds}\cdot\frac{ds}{dt}=\frac{-2\theta}{(2\rho+\theta)^2}\frac{d\rho}{ds}\cdot\frac{ds}{dt}. \tag{1}$$

It is apparent that an actually discontinuous change in the radius of curvature will produce an infinite strain rate, and some testing instruments may approach this.

In order to generate curves along which $(d\epsilon/ds)$ may be expressed as a function of distance along the curve s, it is necessary to obtain ρ as a function of s as well as the representation of the bending curve with s as a parameter. This may be done by employing the familiar formulae for the radius of curvature and for the distance along a rectifiable curve. A more general approach would be to impose a desired function of s in place of $(d\epsilon/ds)$ and to integrate (1). This would yield a general solution in the form of curves along which the strain rate would not only be known, but predetermined for a given velocity of the specimen. When this is done, the following parametric representation for x and y of the bending curve is obtained:

$$y=\int \sin\varphi_1(s)ds+C_1, \tag{2}$$

$$x=\int \cos\varphi_1(s)+C_2, \tag{3}$$

$$\varphi_1=\int\frac{ds}{\varphi_2(s)}+C_3, \tag{4}$$

$$\varphi_2(s)=\frac{\theta}{2}\left\{\frac{1}{C_4}+\int\varphi_3(s)ds-1\right\}, \tag{5}$$

$\varphi_3(s)=\dfrac{d\epsilon}{ds}$, i.e., the function originally selected to give the rate of strain. (6)

It may be noted that φ_1 is identifiable with $\tan^{-1}(dy/dx)$ and φ_2 with the radius of curvature.

This method depends upon four consecutive integrations which may present some difficulty in its application. Accordingly, a more restricted method is now presented which will yield easily describable strain rates. Suppose $y=f(x)$ is a suitable curve along which the distance s may

be measured. Then

$$s=\int\{1+f'(x)^2\}^{\frac{1}{2}}dx \quad \text{and} \quad y=\int f'(x)dx. \tag{7}$$

The problem then reduces to guessing a value of $f'(x)$ so that both integrals may be integrated and so that the integral giving s should yield a function sufficiently simple so that it may be solved in terms of s alone. This will give the essential requirements for calculating $(d\epsilon/ds)$ according to Eq. (1). For example, let $f'(x)=\text{ctn}x$, in which case $y=\log\sin x+c_1$ and $s=\log\tan(x/2)+c$. Furthermore, $\rho=-\cosh s$ and $(d\epsilon/ds)=[2\sinh s/(\theta-2\cosh s)^2]$.

In the application of the analysis to an experiment, certain limitations must be recognized. In the first place, there is a limit to the minimum value that the radius of curvature may assume for a given thickness of specimen, since for insufficiently large values of the radius the bending may not obey the expression given for the tensile strain. Further, in any practical consideration there is an effective upper limit to the velocity of the test piece—hence, an effective upper limit with respect to the strain rate. These limiting factors can only be established by experiment. The method herein described is admittedly an hypothesis, since no experimental work has been done on a device incorporating any of the design features noted above.

Finally, it might be in order to state that the ultimate purpose of a device so constructed would be to determine the critical strain rate quantitatively. Also, it should be possible by adjustment of the curve form to determine the dependence of the critical strain rate (if indeed it exists as a sharply defined property) upon the strain itself.

The author wishes to express his thanks to Dr. J. N. Wilson and Dr. E. M. Brunner of these laboratories for helpful discussions in connection with these remarks.

[1] R. F. Boyer and R. S. Spencer, *Advances in Colloid Science*, "Second-Order Transition Effects in Rubber and High Polymers" (Interscience Publishers, New York, 1946), Vol. II, p. 1.

[2] a. S. C. Myers, Modern Plastics **20**, 81 (1942); b. *Bend Brittle Test: Plasticizers for Use with Polyvinyl Chloride and Copolymers*, The Resinous Products and Chemicals Company, 1944, p. 15; c. Bimmerman and Keen, Ind. Eng. Chem. (Anal. Ed.) **16**, 588 (1944); d. G.D.E.S. No. 18, 1943 (British Specification); e. A.S.T.M. Designation: D746-44T and D736-43T.

On the Flow of a Perfect Fluid between Two Plates from a Circular Inlet Channel

E. T. Benedikt
North American Aviation, Inc., Inglewood, California
August 23, 1948

We shall assume a system consisting of two infinitely extended parallel plates, S_1, S_2 (see Fig. 1), separated by a distance h. A perfect, incompressible fluid will be assumed to enter through a circular orifice σ of radius r_0 from a long cylindrical tube t and move in the tube and between the two plates with a permanent irrotational motion. Absence of cavitation phenomena or vortex formation around the boundary of σ will be assumed.

In what follows, a system of cylindrical coordinates, r, θ, z, having its origin at the center 0 of σ and its axis perpendicular to the plates, will be used.

Under the conditions specified above, the velocity $\mathbf{v}$ of the fluid can be expressed under the form

$$\mathbf{v}=\text{grad}\varphi. \tag{1}$$

The stream function φ is harmonic and subjected to the boundary conditions,

$$(\partial\varphi/\partial z)_{\substack{z=0\\ r<r_0}}=q, \quad (\partial\varphi/\partial z)_{\substack{z=0\\ r>r_0}}=(\partial\varphi/\partial z)_{z=h}=0. \tag{2}$$

It can easily be seen that these conditions are satisfied (in the region between S_1S_2) by the superposition of the flow issuing from the orifice σ and its images σ_1, $\sigma_2\cdots$; σ_{-1}, $\sigma_{-2}\cdots$ (see Fig. 1) situated at the points $r=0$, $z=2nh$ $(n=\pm1,\ \pm2,\ \cdots)$. Making the assumptions (approximately valid if h/r_0 is not too small) that the velocity field in the inlet tube t and on σ is uniform, we shall have for the potential of the disk σ_n the expression[1]

$$\varphi_n=(2q/\pi)\int_0^\infty(e^{\pm\lambda(z-2nh)}/\lambda)\sin(\lambda r_0)J_0(\lambda r_0)d\lambda, \tag{3}$$

where J_k indicates a Bessel function of the first kind of order k, and the upper or lower sign for λ has to be taken according to whether $n>0$, or $n\leqslant 0$. (This is due to the fact that in the first case the flow occurs downwards, in the latter upwards.)

The stream function will therefore be given by the expression

$$\varphi=\Sigma\varphi_n=(2q/\pi)\int_0^\infty(F(\lambda,z)/\lambda)\sin(\lambda r_0)J_0(\lambda r_0)d\lambda, \tag{4}$$

where

$$F(\lambda,z)=\sum_{n=0}^{\infty}e^{-\lambda(z+2nh)}+\sum_{n=1}^{\infty}e^{\lambda(z-2nh)}$$
$$=2\cosh(\lambda z)/(1-e^{-2h\lambda})-e^{\lambda z}. \tag{5}$$

From (1), (4), and making use of the well-known formula $J_0'(x)=-J_1(x)$, we obtain

$$v_r=\partial\varphi/\partial r=-(2q/\pi)\int_0^\infty F(\lambda,z)\sin(\lambda r_0)J_1(\lambda r)d\lambda,$$

$$v_z=\partial\varphi/\partial z=(2q/\pi)\int_0^\infty[F'(\lambda z)/\lambda]\sin(\lambda r_0)J_0(\lambda r)d\lambda, \tag{6}$$

where because of (5) we have

$$F'(\lambda,z)=2\lambda\sinh(\lambda z)/(1-e^{-2h\lambda})-\lambda e^{\lambda z}.$$

The pressure at any point can be obtained from Bernoulli's equation

$$p(r,z)=H-(\rho/2)[v_r^2(r,z)+v_z^2(r,z)], \tag{7}$$

the constant H being the head of the flow, and ρ the density of the fluid.

The force (of interest in certain applications) exerted upon a circular region of radius R, centering on z and belonging to the surface S_2, will be perpendicular to the latter and of magnitude:

$$Z=(H-p_0)S-Z_1,$$

where p_0 is the pressure outside the system, S the area of S_2, and

$$Z_1=\pi\rho\int_0^R v_r^2(r,h)rdr.$$

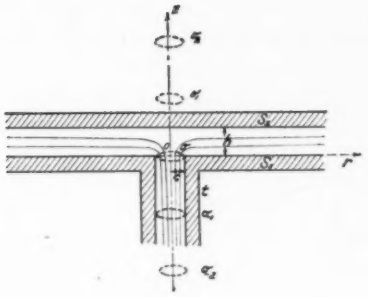


FIG. 1.

Making use of (6) and putting $A=4\rho q^2/\pi$ we shall have

$$Z_1=A\int_0^R\left[\int_0^\infty F(\lambda_1 h)\sin(\lambda r_0)J_1(\lambda r)d\lambda\right]^2 rdr$$

$$=A\int_0^R\int_0^\infty\int_0^\infty F(\lambda_1,h)F(\lambda_2,h)\sin(\lambda_1 r_0)\sin(\lambda_2 r_0)$$
$$\times J_1(\lambda_1 r)J_1(\lambda_2 r)rdrd\lambda_1 d\lambda_2$$

$$=A\int_0^\infty\int_0^\infty F(\lambda_1 h)F(\lambda_2 h)\sin(\lambda_1 r_0)\sin(\lambda_2 r_0)$$
$$\times\left[\int_0^R J_1(\lambda_1 r)J_1(\lambda_2 r)rdr\right]d\lambda_1 d\lambda_2, \quad (8)$$

the interchange of integrations involved in the last step being justified since the integrand satisfies the required conditions of uniform convergence. Applying Lommel's formula for integrals of products of Bessel functions[2] we shall have

$$\int_0^R J_1(\lambda_1 r)J_1(\lambda_2 r)rdr=[R/(\lambda_1^2-\lambda_2^2)]$$
$$\times\{\lambda_2 J_0(\lambda_2 R)J_1(\lambda_1 R)-\lambda_1 J_0(\lambda_1 R)J_1(\lambda_2 R)\}.$$

Making use of this formula and its symmetry with respect to the parameters λ_1, λ_2, formula (8) can be put in the form

$$Z_1=2AR\int_0^\infty\int_0^\infty[\lambda_1/(\lambda_1^2-\lambda_2^2)]F(\lambda_1 h)F(\lambda_2 h)$$
$$\times\sin(\lambda_1 r_0)\sin(\lambda_2 r_0)J_0(\lambda_1 R)J_1(\lambda_1 R)d\lambda_1 d\lambda_2.$$

From formula (5) we obtain, with an easy transformation,

$$F(\lambda,h)=(e^{\lambda h}+e^{-\lambda h})/(1-e^{2\lambda h})-e^{\lambda h}=\mathrm{csh}(\lambda h).$$

Putting, therefore, $\lambda_1 R=\xi$, $\lambda_2 R=\eta$ and, introducing the dimensionless ratios $\alpha=h/R$, $\beta=r_0/R$, formula (9) can be written in the form

$$Z_1=2A\int_0^\infty\int_0^\infty\frac{\xi}{\xi^2-\eta^2}\frac{\sin(\beta\xi)}{\sinh(\alpha\xi)}\frac{\sin(\beta\eta)}{\sinh(\alpha\eta)}J_0(\xi)J_1(\eta)d\xi d\eta.$$

The author wishes to thank Dr. W. Bollay, Dr. S. H. Browne, and Dr. R. P. Isaacs of North American Aviation, Inc., for helpful discussion regarding the problem treated in the present work; and Mr. I. K. Williams of North American Aviation, Inc., for assistance in the preparation of the paper.

[1] See Gray, Matthews, and MacRobert, *A Treatise on Bessel Functions* (London, 1931), p. 141.
[2] See N. W. McLachlan, *Bessel Functions for Engineers* (Oxford University Press, London, 1941), p. 95.

New Booklets

Stanford Research Institute, Stanford, California, founded in February 1947 at the suggestion of a group of West Coast industrial leaders, has made available a 22-page brochure describing the Institute's work and facilities. Free on request.

S. C. Johnson and Son, Inc., Racine, Wisconsin, has issued a 20-page booklet entitled *How to Care for Your Floors*, showing in words and picture how to maintain all types of floors in common use. Free on request.

Eberbach and Son Company, Ann Arbor, Michigan, has published a 4-page leaflet on *Eberbach Laboratory Shakers*. Free on request.

Southwest Research Institute, San Antonio, Texas, has issued a 28-page brochure delineating its functions and services. Free on request to Box 2296, San Antonio.

Allied Radio Corporation, 833 West Jackson Boulevard, Chicago 7, Illinois, has announced the publication of its new 1949, 180-page catalog, covering *Everything in Radio and Electronics*. Free on request.

Here and There

Personnel

Nathaniel C. Fick was recently appointed deputy executive director of the Committee on Basic Physical Sciences, Research and Development Board, with offices in the Pentagon Building, Washington, D. C.

A. M. Zarem, research engineer, has been appointed chairman of physics research and manager of the new Los Angeles Division of the Stanford Research Institute.

Lewis W. Chubb, director emeritus of the Westinghouse Research Laboratories, has retired after 43 years of scientific research with Westinghouse Electric Corporation. He served as director of the Laboratories for 18 years.

Awards

The British Consul General for New York City recently decorated O. S. Duffendack with the ribbon of the King's Medal for Service in the Cause of Freedom for valuable services rendered to the Allied war effort in various fields of scientific research and development. Until 1944 professor of physics at the University of Michigan, Dr. Duffendack is now president, vice chairman of the board, and director of research for Philips Laboratories, Inc.

In September Hornell Hart, professor of sociology at Duke University, was chosen in a nationwide contest among leading social scientists as the winner of the

Edward L. Bernays Atomic Energy Award, offered by the Society for the Psychological Study of Social Issues for the best action-related research in the social implications of atomic energy. The award, a $1000 U. S. Government Bond, was presented to Dr. Hart at the annual convention of the American Psychological Association on September 10.

Physical Society of Pittsburgh

The Physical Society of Pittsburgh, which will observe the 25th anniversary of its organization this year, recently elected the following new officers: E. Stickley, president; R. C. Mason, vice president; E. C. Creutz, secretary-treasurer.

Queens University To Get Synchrotron

General Electric Company's General Engineering and Consulting Laboratory is building for Queens University, Kingston, Canada, a 70-million-volt synchrotron, which will be used for nuclear research. It will be modeled after a machine developed by the G.E. Research Laboratory for its own use.

Airborne Instruments Fellowships

Two fellowships have been established by Airborne Instruments Laboratory, Inc., Mineola, New York, to aid worthy young men to obtain advanced degrees in the field of communications and electronics. To be known as the Airborne Instruments Laboratory Fellowships, one is in the amount of $1000 plus tuition at Stanford University, Palo Alto, California; the other, in the amount of $1200 for a single man or $1800 for a married man, plus tuition, is at the Massachusetts Institute of Technology. Recipients of these fellowships will be selected by the staffs of the electrical engineering departments of the two schools.

Mechanics Colloquium at Illinois Tech

Prominent scientists from the East and Midwest will speak before Illinois Institute of Technology's Mechanics Colloquium during the 1948–49 academic year. First of the eight monthly meetings occurred in October, when R. E. Peterson of Westinghouse Research Laboratory spoke on "Interpretation of service fractures." The Colloquium is open without charge to the public. Persons desiring regular announcements of future lectures should write to Mr. R. L. Janes, Illinois Institute of Technology, Chicago 16, Illinois.

Oak Ridge Radio-Isotope Courses

The Oak Ridge Institute of Nuclear Studies announces that four additional radio-isotope courses have been scheduled for the fall and winter months. The additional courses are being offered because more than 300 qualified individuals applied for entrance to one of three courses held during the summer, and there were places for only 96. Dates for the new courses and the fields of interest which will have priority follow:

October 25, 1948	Industrial personnel
January 3, 1949	Government personnel
February 14, 1949	Medical personnel
March 14, 1949	General

Additional information and application blanks may be obtained from Dr. Ralph T. Overman, Head, Special Training Division, Oak Ridge Institute of Nuclear Studies, P. O. Box 117, Oak Ridge, Tennessee.

X-Ray Laboratory Tribute to William Coolidge

The largest x-ray development laboratory in the world was dedicated on September 13 in Milwaukee, Wisconsin, by the General Electric X-Ray Corporation as a permanent living tribute to Dr. William D. Coolidge, 75-year-old x-ray pioneer and director emeritus of the G.E. Research Laboratory, Schenectady, New York.

Westinghouse Scientists Will Work at Argonne

A scientific "lend lease" program aimed at speeding up all phases of atomic energy research has been initiated between the Argonne National Laboratory, Chicago, and the Westinghouse Electric Corporation. Under the arrangement, Westinghouse will make available to Argonne, at its request, specialists in any required field of research and engineering. The men selected will be granted a leave of absence from their regular duties to work at the Chicago laboratories.